

# Kovarianz oder Bestimmtheitsmaß?

## Didaktische Überlegungen zur Basis der Korrelation

RAPHAEL DIEPGEN, BOCHUM

---

**Zusammenfassung:** Zwei unterrichtliche Wege zum Korrelationskoeffizient  $r$  konkurrieren: erstens der Weg über die Kovarianz, zweitens der Weg über das Bestimmtheitsmaß. Ein Vergleich beider Wege mündet in der These, dass der Weg über das Bestimmtheitsmaß der bessere ist.

Der neue nordrhein-westfälische Gymnasiallehrplan für die Klasse 11 nennt auch „Ausgleichsgerade, Regression, Korrelation“ als obligatorisches Thema im Unterricht über Deskriptive Statistik (Ministerium für Schule und Weiterbildung, Wissenschaft und Forschung des Landes Nordrhein-Westfalen 1999, S. 15). Beschreibende Statistik soll sich also im Unterricht auch auf mehr als eine Dimension oder Variable beziehen, wenngleich augenscheinlich nur im Rahmen linearer Ansätze der klassischen Kleinst-Quadrate-Statistik. Man darf gespannt sein, welche Unterrichtstraditionen sich hier etablieren werden. Im Wesentlichen scheinen sich dafür zwei Kandidaten anzubieten: der Weg zur Korrelation über die Kovarianz einerseits, der Weg zur Korrelation über das Bestimmtheitsmaß andererseits. (Mischformen sind denkbar.) Auch wenn die Etablierung unterrichtlicher Traditionen wohl nur zum geringsten Teil rationaler Diskussion folgt, sei hier dennoch ein Vergleich dieser beiden Kandidaten versucht.

### Der Weg über die Kovarianz

Als „Maß“ für den „Zusammenhang“ zweier Variablen  $X$  und  $Y$ , gestützt auf die jeweils  $n$  Messwerte  $x_i$  und  $y_i$  ( $i = 1, \dots, n$ ), wird zunächst die Kovarianz

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

eingeführt – nicht notwendig in Verbindung mit der Regressionsgeraden – und in folgender Hinsicht diskutiert: Liegen die beiden Werte  $x_i$  und  $y_i$  eines Messwertpaares gleichzeitig beide oberhalb bzw. unterhalb des Mittelwertes der zugehörigen Variablen  $X$  und  $Y$ , so vergrößert dies die Kovarianz;

paart sich indessen die Überdurchschnittlichkeit in der einen Variablen mit der Unterdurchschnittlichkeit in der anderen Variablen, so verringert dies die Kovarianz. Graphisch interpretiert: Teilt man das entsprechende Streudiagramm am Schwerpunkt in vier Quadranten, so vergrößern Punkte in den Quadranten links unten und rechts oben die Kovarianz, während sie von Punkten in den Quadranten links oben und rechts unten verringert wird. Die weitere Erkenntnis schließlich, dass diese Kovarianz als Maß für den Zusammenhang der beiden Variablen  $X$  und  $Y$  noch von den Streuungen

$$s_X = \sqrt{s_X^2} = \sqrt{s_{XX}}$$

und

$$s_Y = \sqrt{s_Y^2} = \sqrt{s_{YY}}$$

in  $X$  und  $Y$  abhängt, führt dann schließlich zu der Produkt-Moment-Korrelation  $r$  als standardisierte Kovarianz:

$$r_{XY} = \frac{s_{XY}}{s_X \cdot s_Y}.$$

Für diese Korrelation ist dann noch gesondert zu beweisen – aus Schülersicht sicherlich recht „trickreich“ –, dass sie betragsmäßig durch 1 limitiert ist.

Offen bleibt bei diesem Ansatz, wieso eigentlich die Kovarianz ein Maß für einen *linearen* Zusammenhang ist; dies wird nämlich aus der skizzierten Begründung für den Schüler kaum deutlich. Diese nämlich bezieht sich nur auf die Lage eines Punktes in den vier durch den Schwerpunkt aufgespannten Quadranten des Streudiagrammes, arbeitet also nur mit den Kategorien „rechts oder links vom Schwerpunkt“ und „oberhalb oder unterhalb vom Schwerpunkt“. Aus Sicht dieser Grobkategorien sind aber die Graphen linearer Funktionen nicht von den Graphen irgendwelcher anderer monotoner Funktionen unterscheidbar, sofern diese auch durch den Schwerpunkt gehen. Und auch grob streuende und durch keine Gerade oder Kurve gut approximierbare Punktwolken müssten zu betragsmäßig hohen Kovarianzen führen, wenn sich dieses grobe Streuen jeweils nur auf zwei einander schräg gegenüber

liegende Quadranten beschränkt. Die skizzierte – offensichtlich um Wesentliches verkürzte – Begründung für die Kovarianz lässt also im Dunkeln, warum die Korrelation  $r$  ein Maß für den linearen Zusammenhang ist, und damit insbesondere auch, warum diese Korrelation betragsmäßig eins wird genau dann, wenn alle Punkte auf einer Geraden (mit von null verschiedener Steigung) liegen, und null genau dann, wenn die Kleinst-Quadrate-Regressionsgerade waagrecht liegt. Ganz grundsätzlich gibt es hier für den Schüler keinen klaren Bezug zwischen der Kovarianz und Korrelation einerseits und der ganz unabhängig davon behandelten oder zu behandelnden Regressionsgeraden andererseits. Will man all diese Argumentationslücken im Unterricht schließen – und darum wird man in einem aufklärerischen Statistikuterricht wohl kaum herumkommen –, so bedarf es dazu einer Fülle recht schwieriger weiterer Überlegungen, deren Komplexität in einem seltsamen Missverhältnis zu der groben Einfachheit der anfänglichen Kovarianzkonzeption steht.

## Der Weg über das Bestimmtheitsmaß

Die Regressionsgerade mit der Gleichung

$$\hat{y}_i = ax_i + b$$

wird bekanntlich so bestimmt, dass die mittels dieser Geraden in Kenntnis der  $X$ -Werte „vorhergesagten“  $Y$ -Werte im Schnitt minimale quadratische Abweichung zu den tatsächlich beobachteten  $Y$ -Werten haben. Die Koeffizienten  $a$  und  $b$  werden also so gewählt, dass die Summe der Abweichungsquadrate

$$SAQ(a, b) = \sum_{i=1}^n (\hat{y}_i - y_i)^2 = \sum_{i=1}^n (ax_i + b - y_i)^2$$

minimal wird. Das Minimum dieser Funktion – nennen wir es  $SAQ_{XY}$  – ist also ein – noch von  $n$  abhängiges – Maß für die Größe der quadratischen Vorhersagefehler bei linearer Vorhersage von  $Y$  aufgrund der Kenntnis von  $X$ . (Der  $n$ -te Teil von  $SAQ_{XY}$  entspricht dem durchschnittlichen quadratischen Vorhersagefehler.)

Stehen die  $X$ -Werte für die Vorhersage von  $Y$  indessen nicht zur Verfügung, ist die im Sinne dieser Kleinst-Quadrate-Philosophie optimale Vorhersagegerade der konstante Mittelwert in  $Y$ ; denn bekanntermaßen ist

$$\sum_{i=1}^n (c - y_i)^2$$

minimal für die Konstante

$$c = \bar{y}.$$

Dieses Minimum, nennen wir es

$$SAQ_Y = \sum_{i=1}^n (\bar{y} - y_i)^2,$$

ist ein – wie oben noch von  $n$  abhängiges – Maß für die Größe der quadratischen Vorhersagefehler bei Vorhersage von  $Y$  ohne Kenntnis von  $X$ . Ist der Unterricht über Regression hier angelangt, ergibt sich sofort die Frage: Um wieviel, also um welchen relativen Anteil, verringert sich der quadratische Vorhersagefehler, wenn man  $X$  für die Vorhersage von  $Y$  berücksichtigt, wenn man also von der „trivialen“ Vorhersage des konstanten Mittelwerts zu einer Vorhersage aufgrund der Regressionsgleichung übergeht – oder graphisch formuliert: wenn man von einer waagrecht zu einer Regressionsgeraden mit beliebiger Steigung übergeht? Einfache Antwort: Um den Anteil

$$r_{XY}^2 = \frac{SAQ_Y - SAQ_{XY}}{SAQ_Y} = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2 - \sum_{i=1}^n (\bar{y} - y_i)^2}{\sum_{i=1}^n (\bar{y} - y_i)^2}.$$

Dies ist das Bestimmtheitsmaß, auch Determinationskoeffizient genannt, dessen mit  $+1$  bzw.  $-1$  multiplizierte Wurzel – je nach positiver oder negativer Steigung der Regressionsgeraden – der Korrelationskoeffizient  $r_{XY}$  ist.

Hier ist dem Schüler sofort klar: Als relativer Anteil liegt  $r_{XY}^2$  – und damit der Betrag von  $r_{XY}$  – zwischen null und eins. Dass  $r_{XY}^2$  insbesondere nicht negativ sein kann – und damit das Wurzelziehen immer möglich ist –, dies ist trivial und bedarf keines komplizierten formalen Beweises. Denn durch die erlaubte Nutzung des Prädiktors  $X$  – graphisch: durch die Erlaubnis, nicht nur den Achsenabschnitt, sondern auch die Steigung der Regressionsgeraden frei wählen zu können – lässt sich die Vorhersage allenfalls verbessern, nicht aber verschlechtern.  $SAQ_{XY}$  ist also trivialerweise allenfalls kleiner, niemals aber größer als  $SAQ_Y$ . Bestimmtheitsmaß  $r_{XY}^2$  und Korrelationskoeffizient  $r_{XY}$  sind eins genau dann, wenn die Regressionsgerade den Vorhersagefehler vollständig, also zu 100%, verringert, wenn also alle Punkte auf einer nichthorizontalen Geraden liegen, und beide Werte sind null genau dann, wenn die Regressionsgerade nicht mehr bringt als die konstante Vorhersage auf den Mittelwert, wenn also die Regressionsgerade waagrecht verläuft – oder anders formuliert: wenn  $X$

für die lineare Vorhersage von  $Y$  nichts bringt. (Weil es keinen dafür nutzbaren linearen Zusammenhang zwischen  $X$  und  $Y$  gibt.)

Durch entsprechende spieltheoretische Rahmung lässt sich  $r^2$  auch ganz anschaulich als Maß für Fehlerkostensparnis interpretieren und erlebbar machen.

Beispiel: Die Schüler erfassen durch Befragen (oder Messen) vieler Probanden  $i$  in einer Urliste deren Größe  $x_i$  und Gewicht  $y_i$ , führen damit eine Regressionsrechnung durch und notieren sich als Ergebnis die üblichen Parameter

$$\bar{x}, \bar{y}, s_x^2, s_y^2, r_{xy}^2$$

sowie die Steigung  $a$  und den Achsenabschnitt  $b$  der Regressionsgeraden. Nach Zufall werden nun in einem Wettspiel vom Spielleiter Probandennummern  $i$  gezogen, und die Schüler müssen auf das Gewicht  $y_i$  des jeweiligen Probanden  $i$  durch Angabe eines Vorhersagewertes wetten, wobei sie letztlich einen Geldbetrag in Höhe der quadratischen Differenz zwischen vorhergesagtem und tatsächlichem Gewicht zahlen müssen. (Beziehungsweise zur besseren Motivation: sich ihr Gewinn um diesen Betrag mindert.) In Spielvariante 1 müssen sie jeweils ohne Kenntnis der Größe  $x_i$  ihre Vorhersage machen (also optimalerweise auf das Durchschnittsgewicht), in Spielvariante 2 wird ihnen jeweils vor Wettabgabe die Größe  $x_i$  genannt, so dass sie optimalerweise mithilfe der Regressionsgleichung ihre Vorhersagen festlegen.  $r^2$  ist dann der relative durchschnittliche Verlustrückgang (auf lange Sicht) beim Übergang von Spielvariante 1 (Durchschnittsverlust  $s_y^2$ ) zu Spielvariante 2. (Natürlich könnte man die Frage nach diesem Verlustrückgang zusätzlich motivieren durch die Frage, wieviel die Schüler denn für die Nennung der Größe  $x_i$  jeweils zu zahlen bereit wären.)

Selbstverständlich taucht auch bei diesem Unterrichtsweg über das sofort interpretierbare und ohne irgendwelche Begründungslücken eingeführte Bestimmtheitsmaß  $r^2$  der Term

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

auf – im wesentlichen also die Kovarianz –, und zwar als Mittelglied bei der Auflösung der zu minimierenden Abweichungsquadratsumme

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

mithilfe der 2. Binomischen Formel. Und selbstverständlich sieht hier der Schüler dann auch, dass die

Kleinheit des Minimums dieser Funktion (und damit letztlich die Vorhersageleistungsfähigkeit der Regressionsgleichung) wesentlich von der Größe dieses Terms, also wesentlich von der Größe der Kovarianz, abhängt. Dies sollte dieser Unterricht auch durchaus betonen, vor allem mit Blick auf die Bedeutung, die Produkte von Zufallsvariablen in der weiteren Stochastik (jedenfalls der Kleinst-Quadrate-Statistik) haben. Nur: Hier ist die Kovarianz ein abgeleiteter, nützlich abkürzender Begriff, nicht aber der Grundbegriff.

## Vergleichendes Fazit

Das Gesagte offenbart bereits die These: Zur Korrelation gelangt man im Unterricht tunlichst über das Bestimmtheitsmaß, nicht über die Kovarianz. Dafür sprechen neben den genannten noch zwei weitere Gründe:

- Die – einzelne – Kovarianz ist ein notwendig auf das Zweidimensionale beschränkter Begriff und insofern grundsätzlich nicht auf mehr als zwei Dimensionen erweiterbar. (Dies gilt selbstverständlich unbeschadet der Tatsache, dass gerade in der multivariaten Statistik typischerweise Varianz-Kovarianz-Matrizen – also jeweils viele Kovarianzen – eine wichtige Rolle spielen.) Demgegenüber ist das Bestimmtheitsmaß – im Statistikerjargon das „multiple  $R^2$ “ – definiert auch für den Fall von mehr als einer Prädiktorvariablen und damit ein sinnvolles Zusammenhangs- oder Modellgütemaß nicht nur für die einfache Regressionsanalyse, sondern auch für weitere Verfahren wie etwa die multiple Regressionsanalyse, die nichtlineare Regressionsanalyse, die Varianzanalyse und die Kovarianzanalyse. Kurzum: Die Kovarianz ist zumindest in diesem Sinne das Spezielle, das Bestimmtheitsmaß das Allgemeine. Schule dient der Allgemein-, nicht der Spezialbildung.
- Das Bestimmtheitsmaß beruht letztlich – anders als die Kovarianz – auf einem Modellvergleich, nämlich auf dem Vergleich des trivialen Vorhersagemodells („konstant der Mittelwert“ oder „Regressionsgerade durch den Schwerpunkt mit Steigung null“) mit dem regressionsanalytischen Modell („linear abhängig vom  $X$ -Wert“ oder „Regressionsgerade durch den Schwerpunkt mit beliebiger Steigung“). Insofern Statistik und Stochastik – wenn nicht gar Mathematik und Wissenschaft überhaupt – wesentlich das Arbeiten mit und in und das Vergleichen von Modellen bedeuten – die Richtlinien und

Lehrpläne werden nicht müde, dies immer wieder als didaktische Leitfigur zu betonen –, erscheint der Weg über das Bestimmtheitsmaß günstiger als der über die Kovarianz. Der Unterricht über das auf deskriptivstatistischer Ebene modellvergleichende Bestimmtheitsmaß mag dann vielleicht auch eine günstige Vorbereitung sein für die spätere Begegnung mit dem entsprechenden inferenzstatistischen Modellvergleich in Form des F-Testes an der Hochschule, dessen stringente mathematische Begründung nämlich nicht nur für Gymnasiasten, sondern auch für Studenten nichtmathematischer Fakultäten zu komplex sein dürfte.

## Literatur

Ministerium für Schule und Weiterbildung, Wissenschaft und Forschung des Landes Nordrhein-Westfalen (1999): Richtlinien und Lehrpläne für die Sekundarstufe II - Gymnasium/Gesamtschule in Nordrhein-Westfalen. Mathematik. Frechen

## Adresse des Autors

Dipl.-Psych. Dr. Raphael Diepgen  
Ruhr-Universität Bochum  
Fakultät für Psychologie

D-44780 Bochum

e-mail: raphael.diepgen@ruhr-uni-bochum.de

---

## Höchst erstaunliche Graphen

HERMAN CALLAERT, DIEPENBEEK, BELGIEN

übersetzt aus Teaching Statistics, volume 22, number 1, S. 25-27, von Nadine Warmuth, Konstanz

---

**Zusammenfassung:** *Erzählt wird eine warnende Geschichte über schlechte Angewohnheiten in graphischen Darstellungen.*

### Einführung

Lernen, zumindest auf dem beschreibenden Niveau, wie man die Information einer Graphik versteht, sollte einfach sein. Tatsächlich haben wir alle eine Menge Erfahrung dadurch, daß wir im täglichen Leben einer Menge „aufsehenerregender“ Darstellungen ausgesetzt sind. Dieser Artikel handelt von einem Graphen, der einige schlechte Eigenschaften hat. Er könnte als Beispiel dafür dienen, die Studenten darauf aufmerksam zu machen, kritisch auf jede Datendarstellung zu schauen, der sie in zukünftiger Arbeit begegnen.

Ob einfach oder kompliziert, Graphen werden überall gezeigt und einige sind weit entfernt von einer optimalen Darstellung. Es ist höchst erstaunlich, wie viele neue graphische Darstellungen den Regeln entsprechen, die in H. Wainers Aufsatz

(1994) formuliert sind (Anmerkung: Es wäre klug, erst den Titel des Aufsatzes zu lesen, bevor man stolz ankündigt, daß zumindest die eigenen Graphen Wainers Regeln entsprechen). Studenten zu lehren, kritisch auf Graphen zu schauen, wird ihnen nicht nur bei der Interpretation, sondern auch bei der Konstruktion von graphischen Darstellungen helfen.

### Ein faszinierender Graph

Arbeitskosten können von verschiedenen Perspektiven her analysiert werden. Eine von vielen interessierenden Größen mißt den „Brutto-Verdienst“. Daten für diese Variable sind erhältlich für Arbeiter in der Industrie und sie werden zusammengefaßt in einer Zahl pro Land der EU. Die Studie bezieht sich auf 1993, was auch erklärt, warum der Verdienst in EUR ausgedrückt wird. Es gibt 16 Datenpunkte, da für den Zweck der Studie „alte“ und „neue“ Länder in Deutschland als separate Objekte behandelt werden. Der zugehörige Graph ist in Abbildung 1 gezeigt.