

KOMMENTIERTE BIBLIOGRAPHIE ZUM THEMA "REGRESSION UND KORRELATION"

von M. Borovcnik, Klagenfurt, und G. König, Karlsruhe

Der am Themengebiet "Regression und Korrelation" interessierte Lehrer (die Lehrerin) möchte, vielleicht angeregt durch die Artikel in den beiden ersten Heften des Jahrgangs 1988 von Stochastik in der Schule, mehr über die mathematische Seite der Konzepte wissen, oder er (sie) möchte sich über die didaktische Literatur zum Thema orientieren. Für diesen Fall bieten wir mit unserer kommentierten Bibliographie eine kleine Hilfe an.

Dabei haben wir im wesentlichen auf die Datenbanken MATHDI (Mathematikdidaktik) sowie MATH (Mathematik) des Fachinformationszentrums Karlsruhe zurückgegriffen. Diese beiden Datenbanken weisen weltweit die Literatur zum Mathematikunterricht bzw. zur Fachwissenschaft Mathematik nach; sie sind gleichzeitig die Online-Versionen der Referateorgane Zentralblatt für Didaktik der Mathematik (ZDM) sowie Zentralblatt für Mathematik (MATH). Nähere Auskünfte erhält man bei G. König, ZDM, 7514 Eggenstein-Leopoldshafen 2.

Wir haben unseren Interessen entsprechend eine Auswahl aus einer Literaturrecherche zum Thema Regression und Korrelation aus beiden Datenbanken getroffen und die dabei erhaltenen Abstracts gemäß der Zielsetzung der Zeitschrift gekürzt, verändert oder auch im Original belassen. Wir danken dem Fachinformationszentrum Karlsruhe für die Erlaubnis dazu und hoffen, dem Leser/der Leserin die eine oder andere Anregung für den Unterricht bieten zu können.

1. Grundsätzliche Diskussionen und methodische Überlegungen

Athen, H.: Regression und Korrelation in der S 2. In: Prax. Math. 23 (1981), S. 106-109.

Die Aufgabe, eine beste Ausgleichsgerade für Datenpunkte zu finden, kann auf verschiedene Weisen in eine mathematische Optimierungsaufgabe umgemünzt werden: Minimieren der vertikalen linearen bzw. quadratischen Abstände zur Ausgleichsgeraden etc. Eingehend beschreibt der Autor die mathematische Herleitung der Regressionsgeraden für den (selten aufgegriffenen) Fall, daß man die Summe der Normalabstände der Punkte von der Geraden minimiert. Die Überlegungen werden vor dem Hintergrund, das Thema Regression und Korrelation im Kursunterricht in der Sekundarstufe II einzubauen, angestellt.

Baumann, R.: Einführung in die beschreibende Statistik unter besonderer Berücksichtigung von Korrelations-, Ausgleichs- und Fehlerrechnung. Grundkurs. Manuskript, 1978, 33 S.

Der Autor möchte Regression und Korrelation wegen der vielfältigen Anwendungen und praxisnahen Fragen in den Grundkurs Mathematik einbauen. Der Stoff ist mathematisch elementar; die zunächst deskriptive Behandlung drängt jedoch von selbst zu anspruchsvolleren Fragen; diese benötigen zu ihrer Diskussion Ideen und Begriffe aus der Wahrscheinlichkeitstheorie. Diese ist nicht Ausgangspunkt unterrichtlicher Überlegungen, wie üblich, sondern dient erst im nachhinein zur Klärung praktisch-statistischer Probleme.

Delanaye, F. und Gregoire, P.: Autor de la determination d'un `meil-

leur` ajustement curvilineare. In: Math. Ped. 49 (1984), S. 81-91.
Für verschiedene nichtlineare Ausgleichskurven werden die Formeln für die Parameter sowie den Korrelationskoeffizienten angegeben. Die Autoren selbst haben ein Programm für die Berechnungen erstellt.

Jost, D.: Elementare Behandlung von Extremwertaufgaben - Anwendung auf die Regressionsgerade. In: Prax. Math. 25 (1983), S. 129-139.
Neben elementargeometrischen Schlüssen und der Benützung von Extrema von Sinus und Cosinus kann folgender Satz zur Behandlung von Extremwertaufgaben ohne Differentialrechnung dienen: "Das Quadrat einer reellen Zahl ist stets nicht-negativ und genau dann Null, wenn diese Zahl selbst Null ist". Man kann damit z.B. ermitteln, welche Konservendose bei gleichem Inhalt minimale Oberfläche hat. Eine geringfügige Modifikation dieses Satzes verhilft dann zur Auffindung der Regressionsgeraden.

Koller, S.: Typisierung korrelativer Zusammenhänge. In: Metrika 6 (1963), S. 65-75.

Die sachliche Deutung gefundener Zusammenhänge liegt außerhalb der statistischen Methodenlehre, obwohl der Praktiker dazu einer genauen Anleitung bedürfte. Der Autor möchte dies mit einer ausführlichen Typisierung korrelativer Zusammenhänge leisten, diese ist gleichzeitig eine Systematik der in der Praxis gängigen Deutungsfehler.

Kroll, W.: Ausgleichsrechnung als Anwendung der elementaren Analysis in Grundkursen. In: MNU 33 (1980), S. 334-341.

Ziel der Ausgleichsrechnung ist es, aus Beobachtungswerten, die unvermeidlich mit Meßfehlern behaftet sind, möglichst gute Näherungen für gesuchte Größen zu finden. Dazu braucht man ein vernünftiges Abweichungsmaß. Diesbezügliche Überlegungen führen zur Methode der kleinsten Quadrate. Ein Weg zur diskreten Minimumsuche für die beste Ausgleichsgerade wird vorgeschlagen, der die Anwendung (mehrdimensionaler) Differentialrechnung überflüssig macht, der aber auch gerade zu dieser Analysis hinführen kann. Der Vorschlag ist methodisch ausgearbeitet und soll den Unterricht um ein anwendungsbezogenes Thema bereichern.

Laußermayer, R.: Korrelation und Kausalität. In: Dörfler, W. und Fischer, R. (Hrsg.): Stochastik im Schulunterricht. Wien/Stuttgart: Hölder-Pichler-Tempsky/B.G. Teubner, 1981, S. 107-116.

Zunächst wird der Korrelationskoeffizient mit möglichst einfachen und schülergemäßen aber doch noch "mathematisch befriedigenden" Mitteln hergeleitet. Sodann wird das grundlegende Problem des Zusammenhangs zwischen Korrelation und Kausalität an zwei Beispielen (in einer für den Unterricht "aufbereiteten" Form) erläutert: (1) Zusammenhang zwischen Pro-Kopf-Einkommen und Energieverbrauch der Weltstaaten. (2) Verursachung von Aggressivität bei Kindern und Jugendlichen durch das Sehen von Filmen.

Puritz, C.W.: Bestimmung von Regressionsgeraden ohne Differentialrechnung. In: Stochastik Schule 2 (1982) 2, S. 29-31.

Für die Ableitung der Regressionsgeraden wird im allgemeinen die partielle Differentiation verwendet. Um dies zu vermeiden, wird in einigen Lehrbüchern eine komplizierte algebraische Methode angegeben. In diesem Beitrag wird dagegen eine elementare algebraische Methode vorgestellt, mit Hilfe derer man die Regressionsgerade bestimmen und die Zerlegung der Varianz in erklärte und nicht-erklärte Varianz herleiten kann.

Wiedling, H.: Beschreibende Statistik. Überlegungen über den Aufbau eines praxisnahen Stochastik-Grundkurses mit geringem Schwierigkeitsgrad. In: MNU 33 (1980), S. 410-415.

Der Autor versucht, den Grundkurs der Studienstufe durch die Regressions- und Korrelationsrechnung praxisnäher zu gestalten. Die Grundidee: "Wie ermittelt man aus Daten eine (lineare) Funktion, die die Abhängigkeit einer beobachteten Variablen y von den übrigen Variablen x_1, \dots, x_n möglichst gut beschreibt?" wird an einem Beispiel ausgearbeitet. Dies führt zur Methode der kleinsten Quadrate, ein Minimierungsproblem, das in ein Gleichungssystem umformuliert wird. Die Güte der äußerlichen Anpassung der Lösung, der Regressionsfunktion, an die Daten wird durch den Bestimmtheitskoeffizienten gemessen. Der Autor geht auf kritische Stellen des Verfahrens ein und gibt methodisch-didaktische Hinweise.

2. Interpretation der Konzepte und Voraussetzungen

Borovcnik, M.: Eine Analyse des x^2 -Unabhängigkeitstests - Möglichkeiten und Grenzen. In: Dörfler, W. und Fischer, R. (Hrsg.): Stochastik im Schulunterricht. Wien/Stuttgart: Hölder-Pichler-Tempsky/B.G. Teubner, 1981, S. 35-48.

Der x^2 -Test auf Unabhängigkeit und der Korrelationskoeffizient werden miteinander verglichen. Die eine Methode läßt ganz andere Fragestellungen beantworten als die jeweils andere, allerdings unter ganz anderen Voraussetzungen. Diese Unterschiede aufzudecken hilft, die Methoden selbst besser zu verstehen.

Dahl, H.: Height, weight, expectation and variance. In: Teach.Stat.4 (1982), S.37-39.

Schüler und Studenten haben häufig Schwierigkeiten mit den Begriffen Erwartungswert, Varianz und Korrelationskoeffizient. Hier wird beschrieben, wie diese Begriffe anhand der Merkmale Körpergröße und Gewicht Schülern leicht zugänglich gemacht werden können.

Franklin, L.A.: Klärung des Konzepts der Regression mit Hilfe von Drei-Punkt-Datensätzen. In: Stochastik Schule 8 (1988) 2.

Zur Wiederholung und zur Vertiefung der Einsicht über die Grundlagen der Regression schlägt der Autor die Beschränkung auf nur 3 Punkte vor. Durch geschickte Wahl der Punkte läßt sich aufgrund der Variation der Daten die Bedeutung der wichtigsten Begriffe wie Varianzen, Residuen und Korrelationskoeffizient klären.

Eisenbach, R. und Falk, R.: Die Abhängigkeit zwischen zwei Größen gemessen als relative Verlustverringerung. In: Stochastik in der Schule 5 (1985) 1, S. 24-31.

Sind zwei Merkmale abhängig, so ist es möglich, aufgrund der Kenntnis des einen Merkmals bessere Voraussagen für das andere zu treffen. Ein Maß für die Abhängigkeit kann nun in der Häufigkeit von Fehlern, die sich durch ein bestimmtes Vorhersageverfahren vermeiden lassen, gefunden werden. Verschiedene Vorhersageverfahren und verschiedene Bewertungen von Fehlern führen zu verschiedenen Abhängigkeitsmaßen. Die diesbezügliche Diskussion kann zur Klärung der Art von Abhängigkeit und deren Erfassung in Kennziffern beitragen.

Knapp, T.R.: Regression toward the mean. In: UMAP J. (1980)4, S.59-71
Regression auf das Mittel ist ein natürliches Nebenprodukt eines

nicht vollständigen Zusammenhangs zweier Variabler, aber Regressionseffekte werden fälschlicherweise oft für Behandlungseffekte in schlecht konzipierten Experimenten gehalten. Dies wird an drei anschaulichen Beispielen gezeigt. Ziel dieser Unterrichtseinheit für die Sekundarstufe 2 ist es, dieses heikle Konzept theoretisch und empirisch zu erklären.

Watkins, Ann E.: The Regression Effect; or, I Always Thought That the Rich Get Richer. In: Math. Teacher. 79 (1986), S. 644-47.

Eine Gruppe von Personen wird zwei Tests unterworfen. Falls die Testergebnisse zufällige Züge widerspiegeln, dann wird folgendes Phänomen (Regressionseffekt) zu beobachten sein: Personen mit den schlechtesten Testergebnissen in Test 1 werden sich in Test 2 im Durchschnitt verbessern, Personen mit den besten Ergebnissen in Test 1 werden sich verschlechtern. Liegt nun zwischen Test 1 und Test 2 ein spezielles Lernprogramm, so mag man versucht sein, den Regressionseffekt ursächlich dem "Erfolg" des Lernprogramms zuzuschreiben. Die Autorin beschreibt diesen Effekt, historische Bezüge und ein klärendes Unterrichtsexperiment.

3. Beispiele für den Unterricht

Coen G.: Langfristige Leistungsprognose im Sportschwimmen. In: Math. Lehren (1984) 4, S. 69-70.

Die Entwicklung der Bestleistungen etwa beim Leistungsschwimmen zeigt langfristig einen Trend. Die dahinterliegenden Gesetzmäßigkeiten zu erkennen und den Trend genau zu bestimmen kann z.B. zu fundierten Vorhersagen über die weitere Entwicklung führen.

Dorn, G.: Die Entwicklung sportlicher Leistungen und die Grenzen der Mathematik. Spekulatives und Nachdenkliches. Unterrichtssequenz für das 11. Schuljahr. In: Math. Lehren (1984) 4, S 62-65.

Am Beispiel der Olympiaergebnisse einiger Schwimmdisziplinen wird die lineare Regression als ein Mittel der Datenanalyse dargestellt. Sie kann Aufschluß über Trends und Entwicklungstendenzen geben, aber auch zu unzulässigen Prognosen und Spekulationen verführen. Zur Bestimmung der Regressionsgeraden wird eine elementare Schwerpunkt-methode verwendet, die die beste Ausgleichsgerade im Sinne der kleinsten Quadrate liefert.

Henningsen, J.: An Activity for Predicting Performances in the 1984 Summer Olympics. In: Math. Teacher. 77 (1984) 5, S. 338-341.

Mit Hilfe der Ausgleichsrechnung wird eine Vorhersage über die Siegerzeiten im 400 m Freistilschwimmen getroffen (mit Arbeitsblatt).

Morris, P.: Eine Facharbeit als Teil der Prüfung. In: Stochastik Schule 5 (1985) 1, S. 17-23.

Ein Sechzehnjähriger untersucht im Rahmen einer Projektarbeit den Zusammenhang von Tabellenplatz einer Profi-Fußball-Mannschaft und ihrer Unterstützung, gemessen an der Zahl der Besucher. Der Schüler wählte das Thema selbst, plante die Untersuchung und zeigte die Schwierigkeiten samt deren Überwindung auf. Das Ergebnis seiner Studie wird vorgestellt.

Pike, D.J.: The Planning of Regression Studies. In: Math. Spectrum. 18 (1985/86) 2, S. 47-54.

Die Beziehung zwischen Alter und Blutdruck wird herangezogen, um zu erklären, wie eine Regressionsgerade gefunden werden kann. Insbesondere wird ausgeführt, wie man geeignete Daten auswählt.

Suich, R. und Rutemiller, H.: Flächen unter Regressionskurven. In: Stochastik Schule 3 (1983) 2, S. 38-45.

Ausgegangen wird vom Problem, die Kohlenmonoxid-Emission zweier PKWs zu testen. Dazu wird eine einfache Erweiterung der auf Quadratsummen-Minimierung basierenden Regressionsgeraden behandelt. Die Diskussion über diese Erweiterung sollte die Schüler und Studenten dazu bringen, gründlicher darüber nachzudenken, wozu die Regressionsgeraden tauglich sind und wozu nicht.

4. Querverbindungen zu Geometrie und Lineare Algebra

Bryant, P.: Geometry, statistics, probability: Variations on a common theme. In: Am. Stat. 38 (1984) S. 38-48.

Geometrische Grundlagen aus der Vektoralgebra und der analytischen Geometrie werden zusammengefaßt und stochastischen Analoga gegenübergestellt. Damit wird die überraschende Kraft einer kleinen Anzahl zugrunde liegender Prinzipien demonstriert. Die Betonung in diesem Beitrag liegt mehr auf der Äquivalenz einiger Begriffe in verschiedenen "Sprachen" als auf der Abhandlung eines Begriffes.

Bücker, R.: Kenngrößen-Bestimmung der Ausgleichsgeraden durch Anwendung linearer Transformationen. In: Prax. Math. 28 (1986) S. 325-326.

Die Ausgleichsgerade durch n Punkte wird nach der Gaußschen Methode der kleinsten Quadrate bestimmt. Im speziellen Fall kann die dazu benötigte partielle Differentiation durch eine lineare Transformation vermieden werden.

Neill, H.: A geometric view of statistics. In: Math. Gaz. 66 (1982), S. 284-294.

An Beispielen wird gezeigt, wie nützlich geometrische Ideen in Verbindung mit Varianz und Regression sind.

Saville, D.J. und Wood, G.R.: A method for teaching statistics using n-dimensional geometry. In: Am. Stat. 40 (1986), S. 205-214.

Es wird beschrieben, wie mit Hilfe der Vektorrechnung Schülern der Sekundarstufe 2 und Studenten im Grundstudium die Themen Varianzanalyse und Regression vermittelt werden können.

Schönwald, H.G.: Geometrische Veranschaulichung von Korrelationskoeffizienten. In: Prax. Math. 24 (1982), S. 202-203.

Faßt man die x- bzw. y-Daten als Vektor \vec{x} bzw. \vec{y} auf, so ist der Korrelationskoeffizient als Cosinus des Winkels zwischen \vec{x} und \vec{y} zu verstehen. Mit Hilfe dieser geometrischen Darstellung kann man einfachere Begriffe der sogenannten Faktorenanalyse besser verstehen.

Sykes, A.: Ein Hoch auf Pythagoras. In: Stochastik Schule 5 (1985) 3, S. 13-16.

Der Nenner (n-1) im Zusammenhang mit der Schätzung der Varianz wird aus geometrischen Beziehungen heraus erklärt. Die verwendeten Bezüge lassen sich später ausbauen, um z.B. die Gemeinsamkeiten von t-Test und Regressionsrechnung zu erklären.

Wolf, U.: Konstruktion der Ausgleichsgeraden dreier Punkte. In: Mathematikunterricht. 28 (1982), S. 57-65.

Die Bestimmung einer Ausgleichsgeraden läßt sich geometrisch umdeuten als Bestimmung des Abstandes eines Punktes von einer Hyperebene. Nach einigen geometrischen Überlegungen wird gezeigt, wie man die Ausgleichsgerade dreier Punkte mit Mitteln der dargestellten Geometrie ohne Rechnung explizit bestimmen kann.

5. Andere Zugänge

Borovcnik, M. und Ossimitz, G.: Materialien zur Beschreibenden Statistik und Explorativen Datenanalyse. Wien/Stuttgart: Hölder-Pichler-Tempsky /B.G. Teubner, 1988.

U.a. werden folgende Techniken behandelt: Inspektion von Punktwolken Anpassen einer Geraden an eine Punktwolke nach Augenmaß; Zerlegung von Punktwolken in (horizontale) Streifen (um Trends besser sichtbar zu machen); Wandernde Kastenschaubilder in den jeweiligen Streifen; Median- und Viertelkurven (Mediankurve z.B. halbiert Punktwolke in den jeweiligen Streifen); Drei-Gruppen-Gerade (Punktwolke wird in drei vertikale Streifen zerlegt, den drei "Medianpunkten" wird eine Gerade angepaßt). Daneben wird ein Vergleich dieser EDA-Techniken mit der Regression nach der Methode der kleinsten Quadrate gezogen.

Haylock, D.W.: Scattergrams and an Elementary Index of Correlation. In: Math. School (1983) 1, S. 37-39.

Folgende Methode zur numerischen Bewertung des Zusammenhangs zwischen zwei Variablen wird an einem Beispiel vorgestellt: Die Punktwolke wird durch je zwei Geraden horizontal bzw. vertikal in neun Quadranten unterteilt. Die vertikalen Geraden werden so gesetzt, daß ein Viertel der Punkte links von der ersten bzw. rechts von der zweiten und die Hälfte der Punkte zwischen den beiden Geraden liegen. Sie markieren gewissermaßen den Boxplot für die x-Daten. Der Korrelationsindex ist nun eine Funktion der Häufigkeiten in den einzelnen Quadranten. Die Methode ist visuell, einfach und an keine Voraussetzungen gebunden.

Griffiths, D.: A pragmatic approach to Spearman's rank correlation coefficient. In: Teach. Statist. 2 (1980), S. 10-13.

Wenn Objekte sich nach mindestens zwei Merkmalen Rangfolge bringen lassen, dann gibt der Spearmansche Rangkorrelationskoeffizient ein quantitatives Maß für die Übereinstimmung der Reihenfolge bzgl. der beiden Merkmale an. Man benützt dazu die Summe der Quadrate der Differenzen der Rangnummern aller Objekte in den beiden Reihenfolgen. Dieser Wert wird auf 1 normiert. Der Beitrag verdeutlicht vor allem den Normierungsprozeß und die Bedeutung solcher Koeffizienten.

Open University: Statistics in Society, MDST 242, Block A: Exploring the data, Unit A3: Relationships. Milton Keynes: The Open University Press, 1984.

Im Rahmen dieses anwendungsbezogenen Statistik-Kurses wird auch ausführlich auf Beziehungen zwischen zwei Variablen eingegangen. Die dargestellten Techniken sind der Explorativen Datenanalyse zuzuordnen: Inspektion von Punktwolken; Anpassen von Geraden zu Punktwolken; Verbessern der Anpassung einer Geraden durch Analyse der Residuen; Robuste Gerade (Drei-Gruppen-Gerade). Die Techniken werden sodann angewendet, um eine makroökonomisch interessante Hypothese

(Phillips-Kurve) über den Zusammenhang zwischen Arbeitslosigkeit und Veränderung der Löhne zu untersuchen.

Wilkie, D.: Pictorial representation of Kendall's rank correlation coefficient. In: Teach. Statist. 2 (1980), S. 76-78.

Eine einfache Methode zur Bestimmung des Kendall'schen Rang-Korrelations-Koeffizienten T besteht darin, daß man zwischen den entsprechenden Punkten der Rangordnung Linien zeichnet; T hängt dann von der Zahl der Überkreuzungen der Linien ab. In Beispielen wird gezeigt, daß die Methode sinnvoll sein kann.

6. Einsatz des Computers

Brandenburg, R.J.: Graphische Auswertung von Meßreihen mit dem Computer. In: Prax. Naturwiss., Phys. 35 (1986) 2, S. 28-30.

Es werden die in der Schule gebräuchlichen graphischen Auswert-Verfahren - Zeichnen von Ausgleichsgeraden, evtl. nach einer Umrechnung der Meßwerte - beschrieben und ihre Übertragung auf den Computer erläutert. Das Verfahren zur Berechnung der Ausgleichsgeraden wird verallgemeinert und zur Berechnung von Ausgleichskurven benutzt. Ein universelles Auswert-Programm für den COMMODORE 64 wird angegeben.

Gale, R.J.: Introductory Linear Regression Programs in Undergraduate Chemistry. In: J. Comp. Math. Sc. Teach. (1982) 5, S. 11-14.

Der Autor präsentiert einfache Programme zur Regressionsrechnung und tritt dafür, Studenten aus technischen Studien diese Programme selbst erstellen anstatt sie fertige Programme benutzen zu lassen.

Hsiao, M.W.: Teaching Regression Analysis with Spreadsheets. In: J. Comp. Math. Sc. Teach. (1985) 3, S. 21-26.

Spreadsheet-Programm-Pakete wie Visicalc, Supercalc oder Lotus ermöglichen auch vielfältige statistische Berechnungen. Der Autor illustriert, wie man multiple Regression mit Hilfe solcher Programmpakete in den Unterricht integrieren kann.

Kawaletz, M.: Ausgleichsrechnung mit dem HP-41C. In: Schumny H. (Hrsg.): Taschenrechner und Mikrocomputer. Jahrbuch 1983. Braunschweig: Vieweg, 1982, S. 39-46.

Der Beitrag behandelt lineare und nichtlineare Ausgleichsrechnung, bespricht die dabei angewandte QR-Zerlegung und gibt einige Beispiele mit Anweisungsliste für den HP-41C.

Reuss, Ch.: Lineare Mehrfach-Regression und Korrelation. In: Mikro- und Kleincomputer (1987) 6, S. 81-83.

Der Autor stellt ein Programm zur Berechnung der Regressionsgeraden vor. Aus einem Satz von bis zu fünf Variablen werden für je zwei Variablen Regressionsgerade und Korrelationskoeffizient berechnet, Punktwolke und Regressionsgerade auf Bildschirm bzw. Plotter gezeichnet. Das Programm berücksichtigt auch Abweichungen in den x-Daten.