Leserbriefe zum Thema „Wie führt man in Regression und Korrelation ein?“

WOLFGANG RIEMER, KÖLN — RAPHAEL DIEPGEN, BOCHUM

Vorbemerkung: *Die Herausgeber von "Stochastik in der Schule" möchten Diskussionen unter den Leserinnen und Lesern dieser Zeitschrift besonders unterstützen. Deshalb bringen wir die Leserbriefe gelegentlich gleich am Anfang des Heftes, vor den Aufsätzen.*
Der Herausgeber

Zum Werkstattbericht eines Schulbuchautors

WOLFGANG RIEMER, KÖLN (04.04.2000)

Bemerkungen zu Raphael Diepgen: Begründungsprobleme im Stochastikunterricht - Werkstattbericht eines Schulbuchautors. (SiS 19(1999) Heft 3, S. 3 f). In der Tat ist es ein reizvolles Unternehmen, einen neuen Lehrplan durch Schulbucharbeit mit Leben zu erfüllen. Auch ich konnte dem Reiz nicht widerstehen und erlaube mir - aus dieser Sicht - einige Anmerkungen zu dem Beitrag von Herrn Diepgen.

1. Partikularinteressen?

Nein, es waren nicht "die Professoren" oder irgendwelche "Partikularinteressen", die der Stochastik (in Form der Korrelations- und Regressionsrechnung) die Tür öffneten, durch welche sie in die Jahrgangsstufe 11 "hineinschlüpfte". Es war wohl eher der Wunsch der Lehrplankommission, auch der Stochastik als der dritten Säule der Schulmathematik eine Chance zu geben, sich dauerhaft und kontinuierlich zu etablieren.

Natürlich wäre es aus der Sicht eines engagierten "Stochastikers" konsequent gewesen, ein Halbjahr Stochastik und eine Abituraufgabe aus diesem Bereich zu fordern. Aber die Kunst einer Lehrplankommission besteht wohl darin, das von der Sache her "optimale" mit dem praktisch Machbaren zu harmonisieren und für das Endprodukt - unseren gültigen Lehrplan - trotz einiger kritischer Stimmen eine breite Akzeptanz im Lande zu sichern. Und das scheint mir mit der Korrelations- und Regressionsrechnung, einem für NRW neuen Gebiet, durchaus gelungen. Dieses Gebiet besitzt schließlich einerseits eine gewisse Tradition ([1], [4] und kann andererseits auch von Lehrern, die selber noch "dazulernen" müssen, recht gut unterrichtet werden. Im Übrigen war die Lehrplankommission ständig im Gespräch mit Autoren und Beratern aller Schulbuchverlage, man vergleiche auch [2].

2. Zur Verknüpfung von Stochastik mit Analysis und linearer Geometrie

Niemand hindert einen Lehrer der Stufe 11 daran, die Regressionsrechnung nach der Analysis in Angriff zu nehmen, wenn man zur Minimierung von Abstandsquadraten die Differenzialrechnung nutzen möchte.

Warum sollte man die in der Regressionsrechnung erarbeiteten Zusammenhänge nicht im Rahmen der linearen Algebra (Stufe 12) aus der Perspektive orthogonaler Projektionen nochmals durchleuchten dürfen? Mit einer *Einführung* der Regressionsgeraden über die Idee der orthogonalen Projektion des R^n auf einen zweidimensionalen Unterraum erzeugte ich bei einem Unterrichtsversuch in einem Leistungskurs eher Unverständnis, nach meiner Erfahrung bedarf es hierzu zweier Curriculumschleifen.

3. Regression und Korrelation - keine Brücke zwischen der Stochastik in SI und SII?

Wir haben in NRW durch den SI-Lehrplan eine recht gute Grundlage für die Stochastik in der Jahrgangsstufe 11. Schülerinnen und Schüler sollten nach einer kurzen Erinnerung in der Lage sein, Wahrscheinlichkeiten (Modellebene) von relativen Häufigkeiten (Realitätsebene) zu unterscheiden. Sie sollten den Zusammenhang zwischen relativen Häufigkeiten und Wahrscheinlichkeiten herstellen können: Wahrscheinlichkeiten als Prognosen für zu erwartende relative Häufigkeiten, relative Häufigkeiten als Schätzungen für Wahrscheinlichkeiten. In den Jahrgangsstufen 9 und 10 konnten sie sich mit Aspekten elementarer Entscheidungsprobleme auseinandersetzen.

Nach meiner Lesart des Lehrplanes, der ja die Idee der Modellbildung sehr groß schreibt und einen Brückenschlag von der SI in die Jahrgangsstufen 12 und 13 *bewusst anstrebt*, ist es nicht nur statthaft, sondern *sehr erwünscht*, in einem Lehrbuch Verknüpfungen zu diesen Vorkenntnissen *anzubieten* - und dem Mittelwert $\overline{X}$ der relativen Häufigkeitsverteilung den Erwartungswert μ der zugehörigen Wahrscheinlichkeitsverteilung, der Stichprobenstandardabweichung s die Standardabweichung σ der zugehörigen Wahrscheinlichkeitsverteilung gegenüberzustellen. Sollte einem Schüler die Wahrscheinlichkeitsverteilung eines Würfels nicht mehr geläufig sein, so ist eine Erinnerung hieran eine Sache weniger Minuten. Aber für den Lehrer, der diese Blickpunkterweiterung wählt, weitet sich die Perspektive. Man *kann* lineare Modelle ($Y= \alpha X+ß + D$, wobei D eine von X unabhängige Störgröße ist) mit Würfeln oder Zufallszahlen simulieren und studieren, wie sich die empirischen Regressions- und Korrelationskoeffizienten den durch das Modell vorgegebenen Werten annähern. Man kann experimentell (Tabellenkalkulation) oder auch theoretisch untersuchen, wie sich die vorgegebene Steigung α und auch die "Störgröße" D auf den sonst so schwer zu interpretierenden Korrelationskoeffizienten auswirkt. [3, S. 75 Aufg. 14, S. 79 Aufg. 7d].

Simulationen, bei denen die Modellparameter zunächst geheim gehalten werden, sich im Verlauf von Versuchen schrittweise immer genauer zu erkennen geben, sind nicht nur spannend, weil man sie "life" im Klassenraum "erleben" kann. Sie machen transparent, wie in der Mathematik - und insbesondere in der Stochastik - Modellebene und Wirklichkeitsebene zusammenhängen.

Wenn man sich entschließt, Wahrscheinlichkeiten einfließen zu lassen, wird auch der "Nenner n-I" zumindest an elementaren Beispielen nachvollziehbar [3, S. 57]. Die "wahrscheinlichkeitsfreie" Begründung von Herrn Diepgen über "Abstände der Messwerte voneinander" ist in meinen Augen sehr originell, Kompliment! Hat uns hier der neue Lehrplan nicht zu einer interessanten fachdidaktischen Innovation verholfen?

4. Experimente - ein frommer Wunsch?

Nein, man muss sich nicht auf Simulationen, wie in Abschnitt 3 erwähnt, beschränken. Warum sollte man nicht den Zusammenhang zwischen Schuhgrößen und Schuhlängen, zwischen Bundweite und Hosengröße, zwischen Rollweite und Körpergewicht beim Ausrollen eines Fahrrades, zwischen subjektiver Sicherheit und tatsächlicher Trefferquote beim Heraushören von Tongeschlechtern u.s.w. ... experimentell untersuchen dürfen? Ich wette, an die Mathestunde, in der jeder seine Schuhe mit Geodreieck (Außenlänge) oder Stahlbandmaß (Innenlänge) oder bei Kartoffeln "Bauchumfang" und Gewicht gemessen und miteinander in Beziehung gesetzt hat, wird man sich auch lange nach dem Abitur noch erinnern.

Genauso spannend dürfte es sein darüber nachzudenken, *wie genau* man aus der Zeugnisnote in der Klasse 5 die Abiturnote oder aus der Größe eines Vorschulkindes dessen Alter vorhersagen kann. Reale Datensätze zu diesen Fragestellungen gibt es in Hülle und Fülle, und wenn man eine Tabellenkalkulation benutzt, sind sie einfach zu handhaben. (Der Anteil der Schülerinnen und Schüler, die zu Hause den Zugang hierzu haben, liegt am Gymnasium heute schon bei 80%!).

Fragen nach der Genauigkeit der Vorhersagen lassen sich beantworten, wenn man die Standardabweichung mit der 68%-Faustregel in Verbindung bringt. Eine "Begründung" dieser Regel ist natürlich in der 11 nur experimentell möglich, aber die in den Klassen 9 und 10 geleisteten Vorarbeiten ($1/\sqrt{n}$-Gesetz) können unterstützend einfließen. Und auch hier lädt die Stochastik in den Stufe 12 oder 13 zu einer vertiefenden, Rückschau haltenden Curriculumschleife ein. Wirklich begründen kann man den zentralen Grenzwertsatz schließlich erst im Hauptstudium mit der Fourier-Transformation oder Methoden der Funktionalanalysis. Aber sollte man deswegen solange auf die Bekanntschaft mit der Normalverteilung verzichten?

5. Der Regressionskoeffizient - Probleme mit dem quadratischen Ansatz ?

Niemand wird den Mittelwert in der Orientierungsstufe über die Minimierung von Quadratsummen "begründen" wollen. Wer zwingt einen, diesen Ansatz zur Begründung des Regressionskoeffizienten einzusetzen? Im Gegensatz zu Herrn Diepgen bin ich der Meinung, dass eine Besinnung auf die bei Mittelwerten gepflegte Ausgleichsidee die Bedeutung des Regressionskoeffizienten sehr transparent hervortreten lässt [3, S.62-63]. Eine weitere, durch den Lehrplan initiierte fachdidaktische Innovation? Der "Kummer" mit dem nicht befriedigend begründbaren quadratischen Ansatz verliert dann seine Bedeutung. Die Minimalitätseigenschaft ist im Nachhinein ein "schönes Geschenk". Und wenn man die Varianz des Merkmals Y gemäß dem linearen Modell zerlegt in $V_Y = \alpha\ V_X+V_D$ erhält man einerseits einen überzeugenden Zugang zum Korrelationskoeffizienten, andererseits wird die in der Stufe 12/13 benötigte Additivität der Varianz vorbereitet ($\alpha = 1$).

Kurz zusammengefasst, bin ich der Meinung, dass der Lehrplan die Offenheit lässt, Korrelations- und Regressionsrechnung nicht nur spannend und realitätsbezogen zu unterrichten, sondern auch so, dass eine Brücke zwischen der Stochastik der SI und den Jahrgangsstufen 12/13 hergestellt wird. Man ist nicht gezwungen, allzu lange in den "anspruchsloseren" Gefilden der Mittelwerte und Diagrammtypen zu verweilen. Im übrigen ist der bei Cornelsen erschienene Grundkurs mit Sicherheit geeignet, trotz der einen oder anderen "schmerzlichen Begründungslücke" mit seinen vielen Beispielen Freude an der Stochastik zu wecken. Die Autoren bei Schroedel und Klett haben sicher nicht weniger "geschwitzt". Jeder Band hat seine Stärken und Schwächen. Der Lehrer, der sich in allen drei Lehrwerken umgeschaut hat, ist für einen fruchtbaren Stochastik - Kurs der Jahrgangsstufe 11 sicher gut gerüstet.

Literatur

[1] Leistungskurs Stochastik. Hannover; Schroedel 1987, ISBN 3-507-83093-0

[2] B. Ralle: Wie Lehrpläne entstehen (können). MNU 52/6 (1999), 323

[3] Lambacher-Schweizer 11 NRW, Stuttgart, Klett 2000, ISBN 3-12-732210-0

[4] Stochastik in der Schule 1/1988. Das ganze Heft ist dem Thema Korrelation/Regression gewidmet

Dr. Wolfgang Riemer
August-Bebel-Str. 80
50259 Pulheim
w.riemer@t-online.de

Zum Leserbrief von Wolfgang Riemer

RAPHAEL DIEPGEN, BOCHUM (25.04.2000)

Dass der im System Erfolgreiche das Glas halb voll nennen muss, das der aus und von dem System ausgeschiedene Kritiker halb leer sieht - nun ja, dies bedarf wohl keiner ausführlichen Stellungnahme.

Stellung nehmen will ich lediglich zu einem mir wichtigen inhaltlichen Punkt, der Behauptung Riemers unter Punkt 5 nämlich, der Unterricht über die Regression könne und solle seinen Ausgang nehmen von der Ausgleichseigenschaft, nicht von der Minimalitätseigenschaft der Regressionsgeraden. Die zweite Eigenschaft ergebe sich im Unterricht sinnvoll im Nachhinein als "schönes Geschenk" der ersten. Wie sieht das bei Riemer aus? Schauen wir uns seinen Einstieg ins Thema Regressionsgerade an (Baum, Riemer u.a., 2000, S. 62):

"In vielen Anwendungssituationen scheint es gerechtfertigt, zwischen zwei Merkmalen X und Y, wie z.B. Körpergröße und Körpergewicht (zumindest in gewissen Bereichen), eine lineare Beziehung $Y=\alpha X+ß$ zu unterstellen. Mit dieser Formel beschreibt man, welchen Wert man für das Merkmal Y erwartet, wenn der Wert des Merkmals X bekannt ist. So stammt die 'Normalgewichtsbeziehung' $Y=1 \cdot X-100$ zwischen Körpergröße X (in cm) und Körpergewicht Y (in kg) von dem französischen Arzt Broca (1824-1880). Hier ist ($\alpha=1$ und $ß= -100$. Für eine Person mit der Körpergröße 172 cm erwartet man nach Broca das Körpergewicht 72 kg.

Natürlich sind gleich große Menschen in der Regel nicht gleich schwer. Daher denkt man sich in jedem Einzelfall das tatsächliche Gewicht Y zusammengesetzt aus dem Normalgewicht $\alpha X+ß$ und einer zufälligen Störgröße D, die bei 'Übergewichtigen' positiv, bei 'Untergewichtigen' negativ ist. Eike (172; 65,2) hat Untergewicht: $D = -6{,}8$ (kg). Sven (185; 90) hat Übergewicht: $D = +5$ (kg). Über die Gesamtbevölkerung gemittelt sollte aber gelten:

$$\bar{d} \approx 0 \text{ bzw. } \mu_D = 0$$

Andernfalls wäre die Bezeichnung 'Normalgewicht' nicht gerechtfertigt. ... In der Regel sind die Parameter α und ß unbekannt, man muss sie aus Messwerten schätzen. Das sei am Beispiel von Erdbeeren erläutert, für die man ebenso wie für Menschen eine 'Normalgewichtsformel' aufstellen kann . ... "

Aus diesem Ansatz wird dann unter der weiteren mehr oder minder stillschweigenden, weil nicht reflektierten oder gar begründeten - Voraussetzung, die Regressionsgerade gehe durch den "Mittelpunkt" der Punktwolke, die übliche Formel für die Regressionsgerade abgeleitet. Erst in einem weiteren Kapitel wird dann gezeigt, dass diese Regressionsgerade den Mittelwert der "vertikalen Abstandsquadrate" minimiert, also in diesem Sinne gut ist.

Warum - so wird der das Mitdenken noch nicht verlernt habende Schüler sofort fragen - ist es in vielen (in welchen?) Anwendungssituationen gerechtfertigt, den Erwartungswert für die Störgröße D als null anzunehmen? Antwort: Keine.

Warum ist diese Annahme wenigstens für die Beziehung zwischen Körpergröße und Körpergewicht bei Mensch und Erdbeere gerechtfertigt? Antwort: Keine. Stattdessen der bloß rhetorisch-zirkuläre Hinweis, andernfalls wäre ja die Bezeichnung "Normalgewicht" sinnlos.

Wie bloß ist Herr Broca zu seiner Formel gekommen? Antwort: Keine.

Gilt für seine Formel $Y=1 \cdot X-100 + D$ überhaupt $\mu_D=0$? Vermutliche Antwort: Wohl kaum.

Und falls doch: Wie ist Herr Broca darauf gekommen? Hat Herr Broca etwa jeweils viele Menschen der Größe 150 cm, der Größe 151 cm, der Größe 152 cm, ... untersucht und jeweils für jede Gruppe das Durchschnittsgewicht bestimmt und dann festgestellt, dass all diese Durchschnittsgewichte der Formel $1 \cdot X-100$ genügen - oder einer ähnlichen Formel $\alpha X+ß$? (An eine entsprechende Empirie scheint der Riemersche Rückgriff auf das Alltagskonzept von "Übergewichtigen" und "Untergewichtigen" anknüpfen zu wollen.) Falls ja, wäre man zwar der Aussage $\mu_D=0$ recht sicher, aber es gäbe überhaupt keine Parameter α und ß mehr zu bestimmen, das regressionsanalytische Problem hätte sich in Luft aufgelöst.

Dürfte man dann berechtigterweise davon ausgehen, dass das bei Menschen empirisch bestätigte lineare Modell $Y=\alpha X+ß+D$ mit $\mu_D=0$ auch für Erdbeeren gilt? Und gar auch noch für all die Merkmale, die dann von den Schülern in den bei Riemer folgenden Übungsaufgaben miteinander korreliert werden sollen, etwa Deutschnote X und Mathematiknote Y? Sollen und können wir uns tatsächlich die Mathematiknote Y als lineare Funktion der Deutschnote X vorstellen, ergänzt lediglich durch eine unsystematische Störvariable D mit Erwartungswert 0? (Vielleicht könnte man dann konsequenter- wie bequemerweise auch auf die Klausuren im Fach Mathematik verzichten und sich auf die Klausuren im Fach Deutsch und einen "störenden" Würfel beschränken. Riemer ist nicht nur schüler-, sondern auch mathematiklehrerfreundlich.)

Kurzum: Die Voraussetzung sich zu Null ausmittelnder Fehler im linearen Modell Y=αX+ß+D lässt sich allgemein ebenso wenig empirisch begründen wie die Voraussetzung, die durch die lineare Beziehung Y=αX+ß beschriebene Gerade gehe durch den Mittelpunkt der Datenwolke. Beide Voraussetzungen lassen sich - wenn überhaupt - nur in ganz bestimmten Anwendungsfeldern (vornehmlich in den klassischen Naturwissenschaften, kaum aber in den Sozialwissenschaften) mit ziemlich komplexen und für den Schüler der Klasse 11 kaum nachvollziehbaren Überlegungen plausibel machen, beispielsweise dann, wenn für die Messungen in Y normalverteilte Messfehler unterstellt werden können.

Beides sind - schon gar für den Schüler der Klasse 11 - letztlich völlig uneinsichtige und unbegründete willkürliche Setzungen, allenfalls rhetorisch kaschiert. Dies aber verbietet sich meines Erachtens für einen Mathematikunterricht, der sich als Schule vernünftigen Argumentierens versteht. Beide Voraussetzungen sind aber auch kaum normativ zu begründen: Warum nur um Gottes Willen sollte man sich wünschen, dass eine Regressionsgerade durch den Mittelpunkt der Datenwolke geht und die Fehler auf lange Sicht ausmittelt? Beides sind völlig unwichtige Eigenschaften.

Das einzige, was man sich sinnvollerweise - und für jeden Schüler der Klasse 11 vollkommen einsichtig - von einer Regressionsgeraden wünschen kann, ist: Sie soll möglichst gute Vorhersagen von X auf Y erlauben, oder anders formuliert, sie soll sich möglichst gut an die Daten anpassen. Und das heißt einzig und allein: Die absoluten oder quadratischen Modellfehler sollen im Schnitt so klein wie möglich sein - ganz gleich, ob sich dann die Modellfehler auf lange Sicht ausmitteln oder nicht. Dies, und nichts anderes, ist sowohl historisch, als auch systematisch die einzige konzeptionelle Basis für die - lineare wie nichtlineare - Regressionsrechnung.

Dass im Falle der Kleinste-Quadrate-Regression dann die Vorhersagefehler im Schnitt Null sind und - nur bei linearen, nicht bei nichtlinearen Ansätzen - für mittleres X mittleres Y vorhergesagt wird, dies sind mathematisch wenig interessante und praktisch ziemlich irrelevante Nebeneffekte. Diese Nebeneffekte zur fundamentalen Basis machen zu wollen, dieser Vorschlag Riemers stellt die Verhältnisse auf den Kopf.

Natürlich ist Riemers Versuch gut gemeint: Er will es den Schülern "mathematisch" so leicht wie möglich machen. Dazu erfindet er gleichsam eine neue Schulstatistik, die auf einfachem Wege dieselben Produkte liefern soll wie die vermeintlich schwierige Hochschulstatistik. Und tatsächlich liefert diese Schulstatistik ja dann auch dieselben formelhaften Konzepte wie Regressions- und Korrelationskoeffizient. Dummerweise aber muss diese Schulstatistik trotz aller rhetorischen Kaschierung auf die eigentlichen Begründungen verzichten, die diesen statistischen Konzepten in der Hochschulstatistik und ihrer Anwendung überhaupt ihren Sinn verleihen. Mir erscheint dies kaum als sonderlich sinnvoller Statistikunterricht.

Literatur

Baum, M., Riemer, W. u.a. (2000): LS 11 NRW. Stuttgart: Klett

Dr. Raphael Diepgen
Fakultät für Psychologie
Ruhruniversität Bochum
44780 Bochum
raphael.diepgen @ ruhr-uni-bochum.de

Zum Leserbrief von Raphael Diepgen

WOLFGANG RIEMER (15.05.2000)

Es sei mir gestattet, auf die von Herrn Diepgen vorgebrachten Einwände gegen meine Konzeption zur Korrelations- und Regressionsrechnung im Grundkurs 11 zu antworten. Ich verkneife mir "polemische Spitzen".

1. Lineare Modelle (in Schule und Hochschule)

In **Hochschulvorlesungen** basiert "die Korrelationsrechnung" beispielsweise auf der Annahme, die Zufallsgrößen (X;Y) seien zweidimensional normalverteilt. Unter dieser Annahme folgt dann zwischen den Zufallsgrößen X und Y eine lineare Beziehung der Form $Y=\alpha X+ß+D$ mit $\mu(D)=0$. Hierbei sind D, X und Y normalverteilt [2, S. 269]. Die Normalverteilungsannahme benutzt man zur Bestimmung der Schwankungen empirischer Korrelations- und Regressionskoeffizienten (Konfidenzintervalle).

In einem **Grundkurs der Jahrgangsstufe 11** hat die zweidimensionale Normalverteilung keinen Platz, aber das "von der Normalverteilung implizierte" lineare Modell $Y=\alpha X+ß+D$ mit $\mu(D)=0$ ist für Schüler theoretisch wie auch experimentell mit Münze und Würfel (oder Tabellenkalkulation) ohne Probleme durchschaubar [l, S. 78 Nr. 5, 6, 7]. Die Zufallsgrößen D, X, Y können im Unterricht diskret, gleichverteilt, binomialverteilt u. s. w. sein. Die Form der Verteilung ist für das Verständnis des linearen Modells bedeutungslos. Immer geht aber die theoretische Regressionsgerade $y=\alpha x+ß$ durch den "Mittelpunkt" $(\mu_x;\mu_y)$, und die durch D modellierten Fehler mitteln sich (*qua Modellannahme*) aus. Konfidenzintervalle für die empirischen Korrelations- und Regressionskoeffizienten müsste man (aber nicht im GK 11) in Abhängigkeit von den verwendeten Verteilungen für jede Verteilung von X und D berechnen oder durch Simulation abschätzen.

Wichtig ist mir, dass *in diesem Modell auch auf der Schule* die funktionalen Abhängigkeiten zwischen Korrelation, Regression und den beteiligten Varianzen vollständig durchschaubar sind [l, S. 71, S. 75, Nr. 14, S.78 Nr. 5, 6, 7d]. Und wenn ein Schüler sie durchschaut hat, hat er einen grundlegenden "Baustein aus dem Haus der Stochastik" verstanden, keinen "formelhaften", sondern sinnvollen Unterricht genossen.

2. Schätzer des Regressionskoeffizienten

Und nun zu der von Diepgen kritisierten Herleitung einer Formel für den Schätzer des Regressionskoeffizienten. *Wenn* eine Punktwolke gemäß dem obigen linearen Modell mit unbekannten Parametern erzeugt wurde, ist $\alpha:= c_{XY}/V_X$, ein sinnvoller Schätzer für den (unbekannten) theoretischen Regressionskoeffizienten α. Und wenn man das im Schulunterricht (rechnerisch) nachweist und zusätzlich vielleicht durch Simulation experimentell bestätigt, hat man als Lehrer (oder als Schulbuchautor) solide gearbeitet. Ob man zum Nachweis eine Ausgleichseigenschaft, [1, S. 63] oder eine Minimalitätseigenschaft [l, S. 68] verwendet, ist Geschmackssache. Jedenfalls ist ein Buch, das beide Aspekte anbietet, besser als ein Buch, das einen Aspekt unterschlägt (und dazu noch verschweigt, dass man auf der Grundlage linearer Modelle arbeitet).

Ich möchte das "plakativ" vergleichen mit der Herleitung einer Formel für die Scheitelstelle einer Parabel mit $f(x)=x^2+px+q$. Der eine geht von der Symmetrie aus und sucht die Stelle a mit $f(a-x) = f(a+x)$, der andere sucht die Stelle, an der f(x) minimal wird. Für Parabeln sind beide Ansätze äquivalent, Unterschiede zeigen sich erst dann, wenn man die Klasse der Parabeln verlässt.

3. Realitätsbezug

Nun wagen wir mit unseren Modellen den Schritt in die Wirklichkeit (der Kräfte, Gewichte, Geschwindigkeiten, Längen, Testwerte, Intelligenzpunkte u. s. w.), wir "stülpen" die Modelle der Wirklichkeit über - oder wir sehen die Wirklichkeit durch die Brille des Modells. Herr Diepgen verwendet als Dozent an der Hochschule im einfachsten Fall seine "Normalverteilungsbrille", Herr Riemer unterrichtet nur an einem Gymnasium und muss sich mit dem noch einfacheren (universelleren) linearen Modell zufrieden geben. Beide sind sich bewusst, dass das Modell die Wirklichkeit nur unzureichend, meist nur lokal angemessen beschreibt und die Güte des Modells im Falle von Rückschlüssen auf die Realität zu prüfen wäre.

4. Broca-Formel

Wenn ich (wie auch alle anderen GK11- Schulbuchautoren) die Broca-Formel als ein den Schülern bekanntes "in ihrem subjektiven Erfahrungshorizont verankertes" Beispiel für ein lineares Modell benutze, heißt das "in Gottes Namen" doch nicht, dass ich so tue, als sei dieses Modell das Kondensat wissenschaftlicher Forschungsarbeit. Ganz im Ge-

genteil: In dem von mir verfassten Text [l, S. 62, 63, aber bitte ganz lesen!] ist mehrfach und explizit von einer *Annahme* die Rede. Auch die lokale Brauchbarkeit ("Gültigkeit in gewissen Grenzen") wird zweimal explizit angesprochen. Deutlicher als durch das (von Herrn Diepgen nicht erwähnte) Einführungsbeispiel auf der gleichen Seite (Beschleunigung eines PKW in verschiedenen Gängen) kann man dem "mitdenkenden Schüler" den (meist nur) lokalen Charakter linearer Modelle kaum vor Augen führen.

Und *wenn* man diese Modellannahme akzeptiert, also im linearen Modell arbeitet, findet man über die Ausgleichseigenschaft eine transparente (weder zirkuläre noch irgendwas rhetorisch kaschierende) Herleitung, welche die inhaltliche Bedeutung von Zähler und Nenner des empirischen Regressionskoeffizienten prägnant hervortreten lässt und inhaltlich verstehbar macht. (Wenn man über den Minimalitätsansatz startet, fallen c_{XY} im Zähler und V_X im Nenner des Regressionskoeffizienten "vom Himmel", ohne dass Schüler damit tragfähige Vorstellungen verbinden.) Mir ist nicht verständlich, mit welcher Leichtigkeit Diepgen, der doch auch einige Schulklassen unterrichtet hat, solche lernpsychologischen Argumente "vom Tisch fegt".

Kurz: Diepgen argumentiert bei seinen Einwänden gegen die Verwendung der Broca-Formel auf der Realitätsebene. Da der zitierte Text aber in der Modellebene spielt, gehen seine Einwände am Kern der Darstellung vorbei. Das Beispiel der Broca-Formel hätte man, wie in einem Manuskriptentwurf geschehen, wörtlich durch das Beispiel einer Computersimulation [l, S. 78 Nr. 6, 7] ersetzen können.

5. Formelhafte Konzepte

Natürlich kann man auch für Punktwolken, die nicht durch lineare Modelle oder gar zweidimensionale Normalverteilungen beschrieben werden, rechnerisch empirische Regressionsgeraden und Korrelationskoeffizienten berechnen. Nur haben diese dann mit der Realität oft wenig zu tun, für *fundierte* Prognosen sind sie ungeeignet. Das deutlich zu machen, hierfür bei Schülern Sensibilität zu wecken, ist Aufgabe eines jeden Unterrichts. Ich habe diesen Auftrag sehr ernst genommen, wie man beispielsweise in [l, S. 65 Nr. 4, 6d, S. 66 Nr. 7, 8, 9, 10, S. 88 Nr. 11, 12, S. 72 Beispiel 2d, S. 75, Nr. 16, S. 83 Nr. 5] nachlesen kann. Und wenn jemand auf der Schule bei kritischer Verwendung griffiger Formeln gelernt hat, Modell und Realitätsebene zu unterscheiden, nimmt er Substantielles mit ins Leben (bzw. in seine Hochschulausbildung), er hat sinnvollen Stochastikunterricht genossen.

Die Tatsache, dass sich viele Anwendungsprobleme nur unzureichend durch lineare Modelle beschreiben lassen, kann nicht als Argument gegen die Beschäftigung mit linearen Modellen und das geistige Durchdringen der im Modell gültigen Regressions- und Korrelationsformeln gelten. Wer wollte - um im Bild zu bleiben - Parabeln aus dem Curriculum der Jahrgangsstufe 9 streichen, nur weil sich wenige Ausschnitte der Realität global durch Parabeln beschreiben lassen?

Literatur

[1] Lambacher-Schweizer 11 (N RW): Stuttgart, Klett Verlag 2000. Best. Nr. 73221

[2] J. Pfanzagl: Allgemeine Methodenlehre der Statistik

Dr. Wolfgang Riemer
August-Bebel-Str. 80
50259 Pulheim
w.riemer@t-online.de

Liebe Leserinnen und Leser!

Welche Meinung haben sie zu den hier angesprochenen Problemen bzw. Unterrichtsvorschlägen? Welches sind Ihre Erfahrungen mit beschreibender Statistik in Klasse 11? Wir bitten Sie um Ihre kritische Stellungnahme!

Zum Leserbrief von Wolfgang Riemer

RAPHAEL DIEPGEN, BOCHUM

Die Diskussion von (SiS 20(2000) 2, 6-7) wird hier fortgesetzt.

1.

Es geht überhaupt nicht um eine unterschiedliche Bewertung der Leistungsfähigkeit und Begrenztheit Linearer Modelle oder gar des Arbeitens mit Modellen überhaupt. Es geht lediglich darum, wie die Berechnung oder "Schätzung" der Parameter dieser Modelle, konkret also der Steigung und des Achsenabschnitts der Regressionsgeraden im Unterricht zu begründen ist. Da gibt es wohl auch kaum einen Unterschied zwischen dem Statistikunterricht in der gymnasialen Oberstufe und meiner statistischen Hochschullehre (nicht etwa für Mathematikstudenten, sondern) für Psychologiestudenten im Grundstudium. Die zweidimensionale Normalverteilung kommt dort beiderseits nicht vor: mathematisches Niveau und "lernpsychologische" Problematik dürften ähnlich sein. "Lernpsychologische" Probleme dürften aber m.E. kaum ein ernsthafter Grund sein, im Unterricht auf die wesentlichen Begründungen mathematischer Konzepte zu verzichten.

2.

Auch wenn zwei Darstellungsweisen mathematischer Konzepte mathematisch äquivalent sind, bedeutet dies noch lange nicht, dass sie in gleicher Weise didaktisch geeignet sind, dass also ihre didaktische Auswahl "Geschmackssache" ist. Beispiel: Der Mittelwert lässt sich im Unterricht wesentlich über seine Ausgleichseigenschaft einführen, er ist dann jener gleichbleibende Summand, der sich ergibt, wenn man eine gegebene Summe in gleiche Summanden zerlegt, und zwar durch ausgleichende Umverteilung von den großen zu den kleinen Summanden und umgekehrt. Mathematisch äquivalent zu dieser Ausgleichseigenschaft ist nun aber die Minimalitätseigenschaft hinsichtlich der quadratischen Abweichungen; in dieser Perspektive erscheint der Mittelwert als jener Wert, für den die Summe der quadratischen Abweichungen minimal ist.

Nun käme wohl kaum ein Lehrer auf die Idee, angesichts dieser mathematischen Äquivalenz den Mittelwert über seine Minimalitätseigenschaft einzuführen: Diese ist nämlich für das wesentliche Konzept des Mittelwerts, d.h. für die typischen Anwendungssituationen, allenfalls von sekundärer Bedeutung, unbeschadet der Tatsache, dass sich durchaus Situationen denken oder konstruieren lassen - beispielsweise Wettspiele mit quadratischer Verlustfunktion -, in denen der Mittelwert nicht seiner ausgleichenden, sondern einzig seiner die Abweichungsquadrate minimierenden Eigenschaft wegen interessant ist.

3.

Übertragen auf die lineare Regression: Selbst wenn auch hier Ausgleichseigenschaft und Minimalitätseigenschaft bezogen auf die Modellfehler äquivalent wären, wäre es eben nicht "Geschmackssache", welche Eigenschaft man im Unterricht zur Einführung nutzt. (Tatsächlich sind diese beiden Eigenschaften hier nicht äquivalent: Aus der Forderung, dass die Summe der quadratischen Modellfehler minimal sei, folgt zwar die Forderung, dass die Summe der Modellfehler null sei - nicht aber umgekehrt, wie beispielsweise ein kurzer Blick auf um den Schwerpunkt punktsymmetrische Punktwolken lehrt, in denen jede beliebige Gerade durch den Schwerpunkt die Summe der Modellfehler auf null bringt, ohne zugleich die Summe der quadratischen Modellfehler zu minimieren.

Für den umgekehrten Schluss von der Ausgleichseigenschaft auf die Minimalitätseigenschaft benötigt man die zusätzliche Forderung, dass die Kovarianz zwischen Prädiktorvariable X und Residualvariable D null sei - eine offensichtlich für die Schüler zu Beginn der Regressionsrechnung wohl kaum leicht zu verstehende Zusatzforderung.)

4.

Wesentlich ist aber für die typischen Anwendungssituationen der Regressionsrechnung nicht der Wunsch, dass sich die Modellfehler ausgleichen - davon hat man nämlich überhaupt nichts sondern einzig der Wunsch, dass die summierten oder durchschnittlichen absoluten oder quadratischen Modellfehler möglichst klein sind, auf dass also das Modell in dieser Hinsicht möglichst gut sei, d.h. möglichst gute Vorhersagen von X auf Y erlaube, oder - anders formuliert - sich die das Modell repräsentierende Regressionsgerade möglichst gut der empirischen Punktwolke anpasse. Es ist hierbei zunächst völlig gleichgültig, wie und wovon abhängig oder unabhängig die Modellfehler verteilt sind, insbesondere also auch, ob man diese Modellfehler als von X unabhängige (beispielsweise normalverteilte) Zufallsvariable mit Erwartungswert null voraussetzen kann - was man mit guten inhaltlichen Gründen wohl ohnehin nur sehr selten können dürfte, schon gar in Klasse 11 des Gymnasiums.

Bei der Regression ist - anders als beim Mittelwert - zunächst die Minimalitätseigenschaft wichtig, nicht die Ausgleichseigenschaft. Wer also die Regressionsrechnung, auf der Ausgleichseigenschaft aufbaut, verpasst m.E. genau die wesentliche Grundidee; er nimmt eine für die typischen Anwendungssituationen der Regressionsrechnung zunächst irrelevante Eigenschaft zur Basis seines didaktischen Aufbaus. Genau so, als führte er den Mittelwert über die Minimalitätseigenschaft ein.

5.

Wer im Riemerschen Sinne die Regressionsrechnung, - vermeintlich schülerfreundlicher mathematischer Vereinfachung wegen - darauf aufbaut, dass die Modellfehler D unabhängig von X mit dem Erwartungswert 0 verteilt sind, muss die linearen Modelle als - mehr oder minder gute - Abbildungen der Realität interpretieren: Denn anders als über weitreichende Annahmen über die "wahren" Beziehungen zwischen X und Y sowie die Entstehung und Natur von D lässt sich diese Basis kaum begründen.

Wer demgegenüber die Regressionsrechnung lediglich über die viel bescheidenere, rein pragmatische Idee der Vorhersagefehler(quadrat)minimierung aufbaut, der kann sich jedweder Annahme über die "Realität" oder "Natur der Zusammenhänge enthalten: Für ihn ist die Regressionsgerade lediglich ein praktisches Instrument zur möglichst guten Vorhersage von X auf Y - bezogen selbstverständlich auf eine gegebene, hoffentlich "repräsentative" Punktwolke -, dessen praktischer Nutzen sich in der möglichst guten Vorhersageleistung - gemessen etwa durch den Korrelationskoeffizienten - erschöpft, nicht aber in einer wie auch immer gearteten "Realitätsangemessenheit".

Kurz gesagt: Die Minimalitätseigenschaft ist ein vernünftiger und einsichtiger Wunsch, der keiner weiteren Rechtfertigung bedarf, die Ausgleichseigenschaft demgegenüber eine gewagte Hypothese über die Realität, die erst durch empirische oder theoretische Plausibilitäten legitimiert werden kann.

6.

Wer die Regressionsrechnung angemessen auf der Minimalitätseigenschaft aufbaut, der läuft daher auch nicht Gefahr, sein Modell als Abbild der Realität überzuinterpretieren. Sein einziges Ziel ist ja lediglich, in willkürlicher, d.h. lediglich der Einfachheit wegen gewählter, nicht aber etwa "realitätsangemessener" Beschränkung auf eine lineare Vorhersagegleichung möglichst gute Vorhersagen von X auf Y zu machen. Das ist alles. Weiter gehende Implikationen etwa über die stochastische Natur der Größen Y, X und D, über die "wahren" Beziehungen zwischen X und Y, über die Verteilung und Unabhängigkeit der Residualvariable D sind hier zunächst überhaupt nicht gemacht.

Aus der Perspektive dieses Ansatzes etwa ist eine Formulierung, wie die folgende von Riemer problematisch: "Natürlich kann man auch für Punktwolken, die nicht durch lineare Modelle oder gar zweidimensionale Normalverteilungen beschrieben werden, rechnerisch empirische Regressionsgeraden und Korrelationskoeffizienten berechnen. Nur haben diese dann mit der Realität oft wenig zu tun, für fundierte Prognosen sind sie ungeeignet. Das deutlich zu machen. hierfür bei Schülern Sensibilität zu wecken, ist Aufgabe eines jeden Unterrichts." Nicht ganz.

Natürlich lassen sich auch bei beliebigen "nichtlinearen" Punktwolken Prognosen mittels linearer Modelle machen; nur sind diese dann halt - erkennbar an niedrigeren Korrelationskoeffizienten, die ja Maß sind gerade für die Prognoseleistung der Regression - weniger gut als Prognosen in "linearen" Punktwolken; sie sind deswegen aber zunächst nicht weniger fundiert" oder "realitätsangemessen", sondern nur weniger erfolgreich.

7.

Der Riemersche Ansatz ist, dies sei gerne zugestanden, der "stochastischere", insofern er eine stochastische Modellierung der Situation impliziert. Der Minimierungsansatz demgegenüber kommt ohne jedwede stochastische Modellierung aus; ihm geht es nur um möglichst gute numerische Approximation. Nur: Ich bezweifle, ob die im Riemerschen Ansatz notwendige stochastische Modellierung in Klasse 11 wirklich geleistet werden kann.

Schon das notwendige Konzept der Nichtkovarianz zwischen Prädiktor und Fehler dürfte hier kaum adäquat darstellbar sein. Insofern scheint mir die nicht so anspruchsvolle Beschränkung auf die Minimalitätseigenschaft für den Statistikunterricht in Klasse 11 ehrlicher.

8. Last not least:

Der Riemersche Ansatz funktioniert nur für den Spezialfall der klassischen L2-Regression, also jener Regression, die die quadratischen Modellfehler minimiert. Keinen Zugang gibt es hier aber zur L1-Regression, also jener Regression, die die absoluten Modellfehler minimiert. Bei der L1-Regression summieren sich nämlich die Modellfehler im Allgemeinen nicht zu null, die Regressionsgerade geht im Allgemeinen nicht durch den Schwerpunkt der Punktwolke, und die Modellfehler korrelieren im Allgemeinen auch nicht zu null mit den Prädiktorvariablen; auch lassen sich dort Streuungen nicht nach dem beliebten Konzept der Varianzzerlegung additiv in unabhängige Anteile zerlegen.

Wer - und dies sollte im Sinne der Richtlinien wohl jeder Lehrer - die nur historisch erklärbare, kaum aber sachlich überzeugende Dominanz der einseitigen klassischen Kleinst-Quadrate-Statistik im Unterricht überwinden und selbst zum Thema machen will, der muss für die Regressionsrechnung einen Einstieg wählen, der nicht nur zur L2-Regression führt.

Dr. Raphael Diepgen
Fakultät für Psychologie
Ruhruniversität Bochum
44780 Bochum
raphael.diepgen @ ruhr-uni-bochum.de