

Kann man die Statistik vor der Mathematik bewahren ?

von Ray Paton, Department of Computer Science, University of Liverpool, England
Bearbeitung : Bernd Wollring, Universität Münster

Zusammenfassung

Plädiert wird für eine Statistikausbildung in den Naturwissenschaften, die genuin von den wissenschaftlichen Problemen herkommt und nicht für eine solche, die durch bereits vorgegebene Werkzeuge und Techniken bestimmt wird. Die Betonung liegt auf einem bedeutungsbezogenen Verstehen der statistischen Ansätze innerhalb des Kontextes der wissenschaftlichen Fragestellung. Der Beitrag schließt mit einer Anwendung dieser Argumentation auf eine Situation im Biologieunterricht der Jahrgangsstufe 10.

Einführung

Statistische Techniken sind unverzichtbar bei wissenschaftlichen Untersuchungen, die von lernenden Schülern oder Studierenden selbst durchgeführt werden. Dieser Beitrag soll zeigen, daß der Wert der Statistik für die Schüler in den Naturwissenschaften in ihrer Anwendbarkeit als beschreibendes Werkzeug liegt, das aus den wissenschaftlichen Fragestellungen direkt erwächst und nicht aus einer mathematischen Theorie. Beispielsweise ist ein Streudiagramm, das Schwefeldioxidemissionen mit Rußkonzentrationen in einer Stadt vergleicht, geeignet, den Schülern ein genaues Bild der Verteilung zu geben, aus dem wichtige Vorstellungen über die atmosphärische Belastung abzuleiten sind. Statistik kann allerdings in der Naturwissenschaft und in der naturwissenschaftlichen Ausbildung mißbraucht werden, denn den Schülern werden oft statistische Techniken an die Hand gegeben, ohne daß ihre Einschätzung auf der Basis einer statistischen Theorie gesichert ist. Dieses Unkenntnis kann zu einem großen Realitätsverlust hinsichtlich der Bedeutung der Daten führen.

Das Problem der Statistik in den Naturwissenschaften

"(a) Ein besonderes Merkmal unfähiger Naturwissenschaftler ist ihr Glaube an die Wirksamkeit statistischer Formeln, insbesondere dann, wenn sie durch einen Computer abgearbeitet werden. Der Gebrauch dieser Formeln und des Rechners wird dann selbst interpretiert als ein Zeichen wissenschaftlicher Mündigkeit." Sir Peter Medawar (1982, S. 167).

Naturwissenschaftler müssen imstande sein, verschiedene Forschungsmethoden zu beherrschen und die Frage zu stellen: "Sind Zahlen alles, was bei wissenschaftlicher Forschung zählt?" Im hohen Alter gab Sir Peter Medawar einige beißende Kommentare über Naturwissenschaftler, die dem Sammeln numerischer Daten zu viel Aufmerksamkeit widmen. In seinem Aufsatz "Unnatural Science" gibt er einige bestimmte Eigenschaften "unnatürlicher Wissenschaftler" an, wie etwa "der Glaube, daß Messen und Zählen für sich bereits wertvolle Tätigkeiten sind" oder "der Glaube, daß Fakten wichtiger seien als Ideen" ("der ganze mißliche Mischmasch induktiver Forscherei"). Die Einschätzung zu Beginn dieses Abschnittes zeigt einen der Gründe für eine Tendenz zur Numerik bei naturwissenschaftlicher Forschung, und es gibt mindestens zwei Gründe, warum eine solche Tendenz nicht von Vorteil ist. Zunächst besteht die Gefahr, daß Forschung um ein bestimmtes aktuelles statistisches Werkzeug herumgebaut wird, daß gerade en vogue ist, eine Art von Anschlußfähigkeit. Andererseits ist das Verlangen nach Daten dazu angetan, die Forschung durch das Setzen bestimmter Marken zu prägen, so daß das Ziel und die Methode einer Forschung gewechselt oder verändert wird, um eine bestimmte statistische Bearbeitung zu ermöglichen. Zum zweiten ist das Verlangen nach numerischen Daten geeignet, die Bedeutung und den Wert der Daten zu verschleiern, ebenso die Interpretation der Ergebnisse. Hudson führt dazu folgendes an:

"Die meisten Sozialwissenschaftler, die sich auf Lochkarten und Computer verlassen, scheinen in dieser Praxis ihre Kräfte zum Nachdenken aufzugeben, und als ein Ergebnis davon sind ihre Daten nahezu unbrauchbar unteruntersucht oder untersucht in einer murksigen verdrehten Weise (in a clumsy ham-fisted fashion)." Hudson (1972, Fußnote S. 64)

Diese Stellungnahmen zeigen, daß einige Wissenschaftler durchaus die Gefahr sehen, die ent-

steht, wenn Statistik benutzt wird und die Verbindung zwischen den Zahlen und ihrer empirischen Bedeutung nicht klar ist.

Dies gilt nicht nur für die Forschung, sondern paßt ebenso auf Schüler und Studenten, die in einer bestimmten naturwissenschaftlichen Disziplin ausgebildet werden. Der Zugang zu statistischen Softwarepaketen ist recht leicht, und es gibt die Versuchung, eine Datenmenge bestimmten statistischen Analysen auszusetzen, die für den Lernenden von der Konzeption her ganz bedeutungslos sind. Betrachten wir den Fall, in dem Schüler in einem praktischen Projekt der Biologie oder Geographie die Höhen zweier bestimmter Pflanzensorten zu beobachten haben, die auf zwei verschiedenen Teilen eines Feldes wachsen, um zu prüfen, ob es ein besonderen Faktor gibt, der dieses Wachstum signifikant beeinflusst. In diesem Fall wird der Lehrende versuchen, einen T-test oder gegebenenfalls einen Mann-Whitney-U-Test zu verwenden. Der Schüler dagegen wird die statistische Theorie, die diesem Test zugrunde liegt, nicht kennen und so, stark gesprochen, keine grundsätzliche Einsicht darin haben, wie dieser Test arbeitet und was er mit den Daten anstellt. Wenn es dann soweit ist, daß er seine wissenschaftliche Kenntnis über die Daten in Form einer einfachen wissenschaftlichen Theorie darlegen soll, kommt es meist zum Versagen.

Statistik und naturwissenschaftliche Erziehung

Jeder Lehrer der Naturwissenschaften wird es für wesentlich halten, daß Schüler und Studierende der Naturwissenschaften verstehen, daß Experimente innerhalb eines gewissen theoretischen Rahmens konstruiert sind, so daß die Meßergebnisse nicht nur numerischen Wert haben sondern eine ganz bestimmte empirische Bedeutung. So hat zum Beispiel der Begriff der Korrelation Bedeutung hinsichtlich der (theoretischen) Statistik und ebenso Bedeutung für die spezielle wissenschaftliche Disziplin, innerhalb der er gerade verwendet wird. Schüler und Studierende der Naturwissenschaften sollten dazu angeleitet werden, ihre Ideen und Theorien zur Korrelation selbst zu entwickeln, und zwar so, wie sie aus der speziellen naturwissenschaftlichen Fragestellung, mit der sie sich gerade befassen, entstehen.

Theoretische Ansätze zur Beschreibung naturwissenschaftlichen Lernens haben gezeigt, daß Kinder, um in dem Verständnis eines Problems voranzukommen, ein theoretisches Verständ-

nis selbst entwickeln (Karmiloff-Smith und Inhelder, 1977) und Ausubels bedeutungsvolle Paradigmen des Lernens halten uns dazu an, herauszufinden, was ein Lernender jeweils weiß, so daß das Lernen selbst von diesem Punkt ausgehen kann. Ein Problem beim Gebrauch der Statistik in der naturwissenschaftlichen Forschung besteht darin, daß die Lehrer oft nicht einschätzen können, was die Lernenden alles nicht wissen und mit Druck fortfahren, statistisch bedeutungslose Techniken voranzutreiben. Im Gegensatz dazu sollten Lehrer herausfinden, wo die Grenzen der statistischen Kenntnisse bei den Schülern liegen und sich darauf einstellen, wie ihre wissenschaftliche Kenntnis entwickelt werden kann, um statistische Ideen darin aufzunehmen. Dies mag auch bedeuten, daß fortgeschrittene Statistik nicht benutzt wird, aber daß das, was benutzt wird, auch verstanden ist.

Die frühe Geschichte der Statistik kann ebenfalls dazu herangezogen werden, diese Position zu verstärken. Einige der statistischen Methoden im Rahmen der Variationsanalyse Korrelations- und Regressionstechnik entstanden aus empirischen Studien im späten 19. Jahrhundert. Zu diesem Zeitpunkt waren die statistischen Konzepte einfach und konnten sehr klar mit den Daten in Zusammenhang gebracht werden. Sie waren direkt aus den Daten abgeleitet. Die Statistik entstand aus der naturwissenschaftlichen Fragestellung, und so wurden auch die Annahmen direkt auf den Daten gegründet und nicht auf einer theoretischen Verteilungsannahme. So notiert etwa Porter (1986) in einem Kommentar zu Galtons Methode der Korrelation: "Er (Galton) begriff es als erster nicht als eine abstrakte Technik der numerischen Analyse, sondern als ein statistisches Gesetz der Vererbungslehre" (Porter, 1986, S. 270).

In diesem Fall erkennen wir, daß eine statistische Konzeption sehr eng an den wissenschaftlichen Kontext gebunden ist. In der Tat verwendete Galton die Statistik als ein integralen Bestandteil seiner wissenschaftlichen Theorie und nicht als ein methodisches Werkzeug, das außerhalb der speziellen wissenschaftlichen Ideen stand.

Statistische Konzeptionen haben eine eigene Bedeutung und müssen mit dem Verständnis der Schüler und Studenten von konkreten Experimenten verbunden werden, wenn sie von diesen nicht selbst angewandt und artikuliert werden. Beispielsweise benutzen Schüler ab etwa 16 Jahren im Bereich der Biologie erwartungsgemäß das abstrakte Werkzeug der Standardabweichung. Um diese Idee richtig einzuschätzen, sind etliche damit verwandte Definitionen erforder-

derlich, die man ebenfalls verstanden haben muß, wenn man das Konzept vollständig und sinnvoll gebrauchen will. Dieses schließt den Mittelwert, die quadratische Abweichung vom Mittelwert, die Summierung dazu und die Freiheitsgrade ein. Darüberhinaus sind diese Begriffe nicht nur einfach durch Formeln verbunden, sondern haben eigene Bedeutungen von sehr großer konzeptioneller Tragweite. Die konzeptionellen Zusammenhänge zwischen diesen miteinander verbundenen Begriffen zusammen machen die Bedeutung der Standardabweichung aus, und dies ist eine große Bedeutung, die die Studierenden der Naturwissenschaften mit einem "Interface", einem Verbindungsglied zwischen ihrem wissenschaftlichen Kenntnisstand und dem statistischen Werkzeug ausstattet. Wie viel Lehraufwand allerdings ist darauf angesetzt, statistische Methoden beizubringen, die für die Schüler und Studenten keine spezielle Bedeutung haben. Der Leser mag sich die folgende Frage stellen: Ist es möglich, sagen wir für Schüler der Oberstufe, von der deskriptiven Statistik und einfachen Notierungen über Untersuchungen zu den Grundgedanken der parametrischen Statistik und des Hypothesentestens vorzustoßen ?

Wer wird Statistik lehren?

"Als ein ehemaliger Statistiker und Mathematiker bin ich fest davon überzeugt, daß Statistik und die Anwendungen der Mathematik in die Ausbildung zu den Naturwissenschaften integriert sein sollten. Ich gebe zu, daß bis zu einem gewissen Grade Statistik besonders auf einem elementaren Niveau im Studium der Mathematik stets enthalten ist. Für die meisten Leute ist dieser Gegenstand nur dann interessant, wenn er auf wichtige und interessante Probleme anzuwenden ist, ganz besonders als Teil der Ausbildung in den Naturwissenschaften, obgleich er auch einen wesentlichen Aspekt der Allgemeinbildung darstellt" (Sir Walter Bodmer, 1989, S. 16).

Naturwissenschaftler haben ein Gefühl für das Verstehen von Daten und für das theoretische Rahmenwerk, innerhalb dessen die Daten interpretiert werden. Statistiker werden nicht denselben Zugang zu den Daten haben. Ihr theoretischer Rahmen ist mehr darauf ausgelegt, innerhalb einer abstrakten mathematischen Theorie Bedeutung zu haben als innerhalb einer Naturwissenschaft. Dieses Problem ist nicht dadurch zu beseitigen, daß man sich nur mit Problemen der täglichen realen Umwelt befaßt. Wenn die Statistik innerhalb der Mathematik

verbleibt, dann bleibt sie auch ein Werkzeug, das gewissermaßen nach Daten Ausschau hält. Wenn ihr Wert darin besteht, die Daten bedeutungsvoller oder aussagekräftiger zu machen, dann sollte sie innerhalb der Naturwissenschaften angesiedelt sein.

Die Antwort auf die Frage: "Wer soll Statistik lehren?" hängt von der Position innerhalb der Naturwissenschaften und der Mathematikausbildung ab. Der Autor schlägt vor, daß innerhalb der Ausbildung in den Naturwissenschaften auf dem Niveau etwa der Sekundarstufe II statistische Werkzeuge in folgendem Zusammenhang benutzt werden sollten :

1. so, wie sie von Untersuchungsproblemen oder Fragestellungen der Datenanalyse direkt entstehen und
2. so, wie sie mit dem Verständnisstand des Lernenden zu vereinbaren sind.

Dies steht im Gegensatz zum Gebrauch fortgeschrittener Techniken, die die Schüler und Studenten im Unwissen lassen und sicherlich nicht darüber Auskunft geben, welche zugrundeliegenden Annahmen in den Testverfahren stecken und welche Beziehung dies zu der Darstellung der Daten etwa zu einem mathematischen Modell hat. Dies ist kein Plädoyer für eine Art ex post facto Ansatz in der Forschung, sondern mehr die Anforderung, die Statistik und naturwissenschaftliche Konzepte in einem vernünftigen Verhältnis ihrer Bedeutung zueinander zu verwenden.

Eine Fallstudie

Die bisher geführte Diskussion ist irgendwie philosophischer Natur, und ein Lehrer der Naturwissenschaften mag wohl fragen: "Wie sieht dies alles in der Praxis aus?" Das folgende Beispiel soll einige Punkte, die in diesem Artikel angeführt sind, verdeutlichen.

Man denke sich eine Schülerin der Jahrgangsstufe 10 im Fach Biologie, die sich mit einem Kleinprojekt zum Studium der Dauer des Zellzyklus bei der simplen Grünalge Chlorella befaßt. Aus ihrer biologischen Kenntnis weiß sie, daß Chlorella ein photosynthetischer Organismus ist, daß sie durch Zellteilung vermehrt kleinere Tochterzellen produziert, die dann selbst wieder wachsen und sich teilen. Der Zweck der Untersuchung besteht darin, herauszufinden, welche durchschnittliche Zeit die Zellen benötigen, um sich zu teilen. Mit ihrem Hin-

tergrundwissen entschließt sie sich, daß die einfachste ihr zur Verfügung stehende Methode darin besteht, die Zellen bei gleichbleibender Beleuchtung zu kultivieren, in Flaschen mit Nährlösung zu halten und regelmäßig zu schütteln. Um die Größe der Kultur von Chlorella zu verfolgen, entschließt die Schülerin sich, diese je einmal in 24 Stunden über eine Zeit von 5 Tagen hinweg abzuschätzen. Nun kommen wir zu der ersten und sehr wichtigen Wechselwirkung zwischen naturwissenschaftlicher und statistischer Denkweise, der Schätzstrategie. Sie trifft die Entscheidung, die Flasche zu schütteln und eine Menge von Zufallsschätzungen durchzuführen. Der Grund für das Schütteln, das zufällige Auswählen und das Erstellen einer entsprechenden Datenmenge anstelle einer einzelnen Erhebung basiert auf einer Theorie über die Homogenität der Kultur. Dies ist ein sehr wichtiger Punkt. Er unterliegt sehr vielen biologischen Experimenten.

Beim Schätzen der Anzahl der Zellen in der Kultur entschließt sie sich, eine Zellkammer (ein Haemocytometer) zu benutzen, und da die Anzahl der Zellen klein ist, zählt sie alle diese Zellen in dem Haemocytometervolumen. Auf diesem Wege braucht sie keine andere Theorie. Sie braucht dabei nicht zu wissen, wie die Zellen in dem Sichtfeld des Haemocytometers verteilt sind. Die Zeit in der Schulstunde gestattet es ihr, 15 Haemocytometerzählungen durchzuführen. Nach 5 Tagen hat sie eine Tafel von 5 Spalten und 15 Zeilen zusammen (siehe Tabelle 1 auf der folgenden Seite).

Eine weitere wichtige Frage zur Wechselwirkung von naturwissenschaftlichem und statistischem Denken entsteht nun: "Wie ist diese Datenmenge zu analysieren und darzustellen?" Eine einfache Lösung wäre es, die Mittelwerte für jeden Tag zu vergleichen, aber dies würde nichts aussagen über den Bereich der erhaltenen Werte und ihren Zusammenhang zum Mittelwert. Also welcher Mittelwert soll gewählt werden?

Schüler, die sich in der Oberstufenmathematik auskennen, werden wissen, wie beschreibende Statistik auf Datensätze anzuwenden ist. Diese Kenntnis kann hineingenommen werden in einen Ansatz zur Analyse insgesamt ohne die Notwendigkeit, statistische Tests zu erarbeiten. Eine sehr gute zeichnerische Repräsentation, die man benutzen kann, um die Daten eines jeden Tages zusammenzufassen, ist ein Stengel- und Blatt-Diagramm. In diesem Fall ist die Information über den Median, den Bereich, die obere und die untere Quartilgrenze zusam-

Sample Count	Day 1	Day 2	Day 3	Day 4	Day 5
1	17	16	16	18	26
2	15	19	20	17	26
3	13	16	17	16	23
4	14	16	18	23	29
5	16	19	20	16	22
6	15	17	21	19	18
7	12	18	18	22	27
8	14	16	18	20	23
9	13	18	16	19	21
10	14	15	22	25	26
11	15	17	18	21	27
12	16	17	21	19	25
13	13	18	19	21	24
14	14	17	19	25	24
15	15	17	19	21	25

Tabelle 1 : Gesamtzahlen von Chlorella im Haemocytometer
(Volumen 1/400 mm³)

men so dargestellt, wie Bild 1 dies zeigt. Es gilt für Spalte 1 von Tabelle 1.

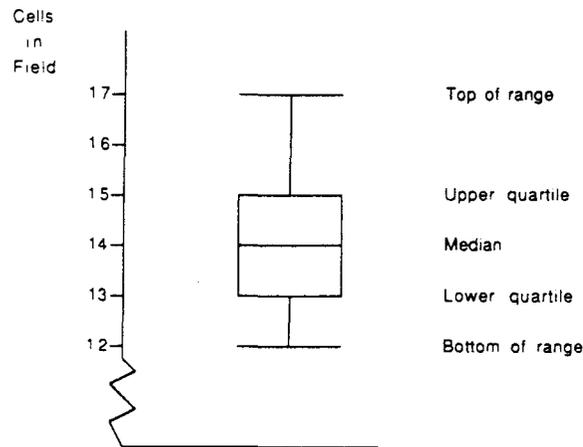


Bild 1 : Stengel-und-Blatt-Diagramm
zu den Daten des ersten Tages

Nun haben wir ein zusammenfassendes Diagramm zu der Datenmenge. Es gibt uns eine

Menge an Informationen, und wenn man es auf die gesamte Datenmenge anwendet, kann es uns helfen, über die Verdoppelungszeit zu befinden. Darüberhinaus kann es im Rahmen der Diskussion in diesem Beitrag dazu dienen, zwei Probleme zu vermeiden. Zunächst die Gefahr der zu geringen Ausnutzung der Daten. Sie wird dadurch vermieden, daß wir hier benötigt sind, die Datenmenge als ganzes anzusehen und in Beziehung zu ihren Teilen. Dieses ist alles in dem einen Stengel-und-Blatt-Diagramm zusammengezogen. Zum zweiten die Gefahr des Überausnutzens der Datenmenge. Sie ist vermieden, weil die Techniken, die hier genutzt sind, direkt aus den eigenen Kenntnissen der Schülerin entstammen, sowohl zu dem wissenschaftlichen Zusammenhang als auch zum Zweck der Untersuchung und auch aus ihrem eigenen Verstehen von Statistik. Der Leser sei aufgefordert, diese Übung mit seinen Schülern mit dem Rest des Datenmaterials aus Tabelle 1 durchzuführen.

Einige weitere Beispiele dazu, wie sich Stengel-und-Blatt-Diagramme in der Jahrgangsstufe 10 sinnvoll verwenden lassen, entnehme man einem angekündigten Beitrag des Autors in Teaching Statistics und Beiträgen zur Explorativen Statistik.

Schlußbemerkung

Der Autor ist der Ansicht, daß es nach wie vor einen Bedarf für einen größeren Dialog zwischen naturwissenschaftlichen Lehrern und Statistiklehrern gibt. Es gibt viele Wege, auf denen das Verständnis des Lernenden für die Zusammenhänge zwischen Naturwissenschaft und Statistik verstärkt werden kann in Situationen, bei denen Daten im Kontext von Theorien sowohl statistischer als auch naturwissenschaftlicher Art zu interpretieren sind.

Literatur

AUSUBEL, D. : Educational Psychology: A Cognitive View .- New York: Holt, Rinehart and Winston, 1968
 BODMER, W. : "Scientific literacy for health and prosperity" .- in School Science Review 70 (253), 1989, p. 16.
 HUDSON, L. : The Cult of the Fact .- London: Jonathan Cape, 1972
 KARMILOFF-SMITH, A. and Inhelder, B. : "If you want to get ahead, get a theory" .- in

Johnson.-Laird, P. and Wason, P.C. : Thinking: Readings in Cognitive Science .-
Cambridge: University Press. 1977, pp. 293-306.

MEDAWAR, P. : Pluto's Republic .- Oxford: University Press. 1982

PATON, R. : "Rank-based Statistics in Biotechnology" .- angekündigt für eine kommende
Ausgabe von Teaching Statistics.

PORTER, T.M. : The Rise of Statistical Thinking 1820-1900 .- Princeton: University Press.
1986