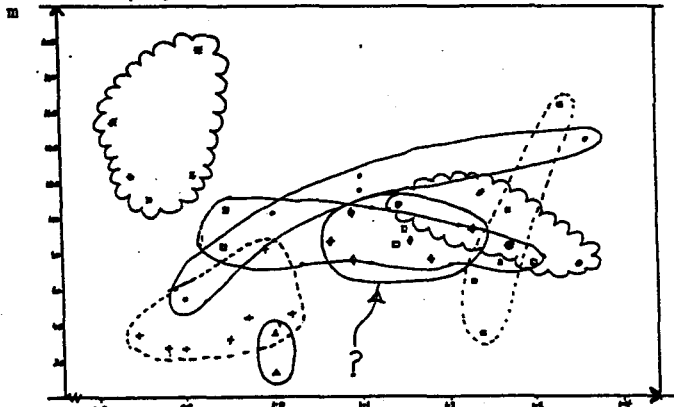


Schüler dann anhand einer Kurzgeschichte den jeweiligen Autor erkennen. Die Daten des englischen Originalartikels sind auch in dieser Darstellung angegeben. Wer macht sich die Mühe den "J.Swift"-Punkt herauszusuchen?

- Ⓐ *Dune* by Frank Herbert
- Ⓑ *The Snow Walker* by Farley Mowatt
- Ⓒ *The Invention of the World* by Jack Hodgins
- Ⓓ *Around the World in Eighty Days* by Jules Verne
- Ⓔ *Split Delaney's Island* by Jack Hodgins
- Ⓜ *Exodus* by Leon Uris
- Ⓝ *For Whom the Bell Tolls* by Ernest Hemmingway
- Ⓟ *Mystery Author*



x = Mittelwert von  $\log_{10}$  (Satzlänge)

Weitere Entwicklung

Eine Vielzahl weiterer Statistiken über den Stil eines Autors können erstellt werden. Unser Projekt soll weiterverfolgt werden; wir wollen dabei auf die folgenden beiden Statistiken Wert legen: Maurice Kendall beschreibt einen "Verschleierungs-Index", der mit der Satzlänge und langen Worten korreliert. Ein Verschleierungs-Index von 10 oder weniger läßt ihn das Buch als gut lesbar empfehlen, während bei einem Verschleierungsindex von 25 und mehr Vorsicht bezüglich der Lesbarkeit angebracht sein soll. Jack Hodgins ist der Meinung, daß der Anteil von Adverbien und Adjektiven in Sätzen ebenfalls sehr viel über den Autor aussagt. Welche Auswirkung dieses Projekt auf die gängige Meinung, daß Mathematiker "nicht literarisch veranlagt" seien, hat, mag die Zukunft zeigen.

Einige Vorstellungen der Studenten über den Median und den Modalwert

von G. V. Barr

frei übersetzt von B. Stumpf

Der Zweck dieses Berichts ist es, in einem Versuch darzulegen, wie Studenten die statistischen Begriffe "Median und Modalwert" verstehen.

Dieser Artikel bezieht sich auf eine Musterstudie, die der Autor durchgeführt hat. Sie versucht folgende Fragen zu beantworten, die mit den studentischen Fehlern zusammenhängen:

- Welche Schwierigkeiten gibt es ?
- Wieviele Studenten haben solche Schwierigkeiten ?
- Wie wird das statistische Vokabular bezüglich dieser Begriffe bekannt ? (entwickelt)

Die Studie betrifft 95 Studenten im Alter zwischen 17 und 21 Jahren, welche alle die mathematischen Grundlagen für ein Studium der Technik mitbrachten. 69 % studierten Ingenieurwissenschaften, die restlichen 31 % Naturwissenschaft und Technik. Von beiden Gruppen wurde erwartet, daß sie Lageparameter und Streuungsparameter aus numerischen Daten berechnen können. Die vorgestellten Beispiele stammen aus den Tests, die dafür entwickelt wurden. Es sind Multiple-Choice-Tests, wobei aus vier Antworten a) b) c) d) die richtige ausgewählt werden soll. e) kann geantwortet werden, wenn das errechnete Ergebnis mit keiner der vier Auswahlantworten übereinstimmt. f) Hier werden diejenigen zusammengefaßt, die keine Antwort geben konnten.

Der Median

Tabelle A gibt die Anzahl der zerbrochenen Stücke in einer Packung Schokolade an.

Anzahl	Häufigkeit
2	5
4	6
6	7
9	2
11	1

Frage: Was ist der Median dieser Verteilung ?

Antworten: a) 4 (20%) b) 5 (33%) c) 7 (14%)  
d) 8 (2%) e) (24%) f) (6%)

Bemerkung: 21 % unter den 24 %, die sich für e) entschieden, haben 6 als Wert angegeben.

Das unterstrichene Ergebnis ist das richtige.

Die häufigste Antwort ist 5. Dies ist der Median der Ziffern in der Häufigkeitsspalte. Die zweithäufigste Antwort ist 6. Dies ist der Median der Anzahlspalte. 20 % der Studenten gaben die richtige Antwort. Weitere 14 % gaben die mittlere Zahl der Häufigkeitsspalte an.

Die Vorstellung, daß der Median Mittelwert von irgend etwas ist, ist klar erfaßt worden. Offensichtlich bestehen nur Zweifel, was man unter dem "irgend etwas" verstehen soll. Ist es etwa schwierig zu verstehen, daß eine Häufigkeitstabelle nur eine Datenzusammenfassung ist, so daß eine andere Darstellung der Daten besser wäre:

2,2,2,2,2,4,4,4,4,4,4,4,6,6,6,6,6,6,6,9,9,11 ?

Sicherlich macht dies einen Teil der Schwierigkeiten der Studenten aus.

Eine andere Möglichkeit wäre, nach dem Median der Anzahl zerbrochener Stücke in den Schokoladepackungen zu fragen. Dies würde für die Studenten das nebulöse Wort "Verteilung" beseitigen. Wäre dies der Fall, würde es auf eine Verständnislücke beim Begriff "Verteilung" hindeuten, so daß ein Student sich fragt: Median wovon ? Etwas ist jedoch bereits gelernt worden: Der Median ist Mittelwert. So wollen wir in noch einem Beispiel untersuchen, was die Studenten sich unter dem Median in einem einfacheren Zusammenhang vorstellen.

Der Median der folgenden Zahlenmenge

1, 5, 1, 6, 1, 6, 8 ist:  
a) 1 (17%) b) 4 (8%) c) 5 (50%)  
d) 6 (21%) e) ( 3%) f) ( 2%)

Wiederum sehen wir, daß die Mehrheit den Median als "Mittelwert" auffaßt. 21 % nehmen den mittleren Wert der Liste. 50 % nehmen den mittleren Wert der geordneten Liste, wie es richtig ist. An dieser Stelle soll die Aufmerksamkeit des Lesers auf den Mangel an Verständnis bei den Studenten für den Begriff des Medians gelenkt werden. Welche Chance haben diese Studenten, solche Vorstellungen außerhalb des Colleges zu benutzen, um zum Beispiel statistische Argumente über Lohnskalen zu verfolgen, wenn nur 50 % der Stichprobe den einfachsten Gebrauch des Medians richtig beherrschen ?

Der Modalwert

Das folgende Beispiel wurde nach der Tabelle A konstruiert.

Frage: Was ist der Modalwert der Verteilung ?

Antworten: a) 11 (4%) b) 7 (13%) c) 6 (68%)  
d) 1 (1%) e) ( 9%) f) ( 4%)

Die Mehrheit scheint die Regel für den Modalwert gelernt zu haben. Es ist der verbreitetste, häufigste Wert und hinsichtlich der Tabelle der Wert, der mit der größten Häufigkeit auftritt. Aber 13 % scheinen diese Regel auf die Häufigkeitsspalte für ihre Antwort angewendet zu haben.

Wie jedoch Studenten den Modalwert in der Gesamtheit statistischer Maße verstehen, ist recht interessant. Betrachten wir dazu folgendes Beispiel:

Welches der folgenden Maße ist nicht zentral (d.h. lokal) ?

a)arith.Mittelwert(5%) b)Standardabweichung(55%) c)Median (5%)  
d)Modalwert (24%) e) ( 1%) f) ( 9%)

Beinahe ein Viertel meint, daß der Modalwert kein zentrales Maß sei. Da Studenten dieser Ausbildungsstufe nur mit lokalen und Streuungsparametern vertraut sind, vermutet man, daß der Modalwert als Streuungsmaß gedeutet wird. Dies ist aber nicht der Fall, wie das Beispiel zeigt:

Welches der folgenden Maße ist ein Streuungsmaß ?

a) Häufigkeit (27%) b)Modalwert ( 0%) c)Histogramm (40%)  
d) Standardabweichung(26%) e) ( 3%) f) ( 3%)

Aus diesem Beispiel können wir erkennen, daß der Modalwert nicht als Streuungsmaß angesehen wird. Aber was wird denn als "Streuungsmaß" angesehen ? 40 % meinen, daß das Histogramm ein Streuungsmaß sei (ein Maß für die Verbreitung). Ein Histogramm zeigt die Verbreitung der Daten, aber es ist sicher keine Maßzahl. Ein ähnliches Argument gilt für die Antwort: Häufigkeit. Die Ausbreitung kann aus einer Häufigkeitsliste ersehen werden - man kann sogar ein Histogramm daraus zeichnen - aber es sind keine Streuungsmaße. Es scheint so, daß die Studenten keine klare Vorstellung von Maßen haben.

Überlegungen und Folgerungen

Oberflächliche Begriffe scheinen im Zusammenhang mit Median und Modalwert erlernt worden zu sein. Die Antworten der Studenten zeigen folgende Regeln: Der Median ist der Mittelwert und der Modalwert ist der häufigste Wert. Darüber hinaus scheinen die Studenten verwirrte Ansichten zu haben, wie man diese Regeln anwenden muß. Warum existiert diese Verwirrung ? Warum fragen sich Studenten: Welche Zahlenmenge wähle ich in dieser Häufigkeitstabelle ? - welche Spalte ?- usw. Was bedeutet der Medianbegriff für

Rang - Statistik

R. Kapadia

Übertreten von Ingeborg Strauß

aus: Teaching Statistics, Bd. 2, Heft 2, 1980, S. 50 f.

sie ? Was bedeutet die Häufigkeitstabelle für sie ? Um den Modalwert zu finden, suchen wir die größte Häufigkeit - aber um den Median zu finden, müssen wir die Tabelle erweitern oder umändern. Dieser Prozeß ist der erste Schritt, um den Median zu finden. Es scheint so, daß der Median schwerer als der Modalwert zu finden ist. Wirkt dies abschreckend ?

Wenn man weiß, wie man eine Häufigkeitstabelle konstruiert, und wenn man Grundvorstellungen über den Median besitzt, könnte man den Median intuitiv aus einer Häufigkeitstabelle bestimmen. Intuition erfordert ein gewisses Einfühlen in die Vorstellungen. Vielleicht fehlt es daran.

Wenn wir über den arithmetischen Mittelwert, den Median und den Modalwert unterrichten, rechnen wir sie oft zu den lokalen Maßen, den Positions(Lage-)maßen oder den zentralen Maßen. Dies scheint auch in der technischen Grundausbildung unterstützt zu werden, indem man den Studenten für diese Begriffe "Aufhänger" gibt. Wir diskutieren die Eigenschaften dieser Maße und entwickeln so ein spezifisches statistisches Vokabular. Dieses Vokabular verkürzt die Wege, wie es eine gute Notation leistet. Das Voranschreiten zu höheren Begriffsbildungen wird schneller. Voranschreiten ist jedoch nur möglich, wenn eine gewisse Vertrautheit mit Grundbegriffen vorhanden ist. Die Verständigung wird schwerer, wenn der Lehrer ein allgemeines Begriffsverständnis voraussetzt, das nicht vorhanden ist. Die Vorstellungen der Studenten entwickeln sich in unzusammenhängender Form, und Verwirrung ist die Folge.

Allgemeine intuitive Vorstellungen über zentrale und Streuungsmaße dürfen nicht die Feinheiten einer genauen Definition verdecken. Zum Beispiel reden wir oft von der Streuung als einem Betrag für die Ausbreitung. Es ist überraschend, daß Studenten verwirrt sind, wenn wir sie zwischen Maßen und Darstellungen für die Ausbreitung unterscheiden lassen. Ihr oberflächliches Verständnis der Grundbegriffe läßt sie zwischen Maß und Darstellung nicht unterscheiden. Sie haben gelernt, daß Streuung zur Ausbreitung gehört genauso wie der Median zum Mittelwert. Dieses oberflächliche Verstehen mit Intuitionen vermischt führt zum Mißverständnis einfacher Probleme und der Grundbegriffe.

Die in diesem Artikel behandelten Probleme beziehen sich auf die statistische Unterweisung der Grundstufe. Sie illustrieren die Feinheiten, die bei den Grundbegriffen vorhanden sind.

Das Interesse der Öffentlichkeit gilt heutzutage u.a. den, so wird vermutet, ständig nachlassenden Leistungsanforderungen in der Schule. Daher ist die Anziehungskraft eines einfachen Schätz-Index, der über die Qualität einer Schule Auskunft gibt, enorm groß. Ein aufstrebender, ehrgeiziger Politiker kann sehr wohl für seine Karriere Punkte sammeln, wenn er sich solch ein - möglichst unkompliziertes - Verfahren für den Leistungsvergleich von Schulen untereinander ausdenkt. Zu den Verlierern gehört dabei die Wahrheit.

Es ist sicher (auch für den deutschen Leser) instruktiv, die Unangemessenheit einer jüngst erstellten 'Hackordnung' von Schulen in West-Mittellengland demonstriert zu bekommen, die eine Rangfolge haargenau festlegt. Ein leitendes Mitglied des Regionalrates präsentierte die folgende Tabelle, in der allerdings aus leicht einsehbaren Gründen die Namen der betroffenen Schulen ausgespart wurden:

Schule	A	B	C	Schule	A	B	C
1	6	166	27,7	16	26	152	5,8
2	22	301	13,7	17	84	481	5,7
3	27	330	12,2	18	13	66	5,1
4	38	429	11,3	19	15	73	4,8
5	89	991	11,1	20	35	144	4,1
6	38	412	10,8	21	98	375	3,8
7	43	446	10,4				
8	49	441	9,0				
9	32	277	8,7				
10	86	722	8,4				
11	61	500	8,19				
12	88	720	8,18				
13	39	287	7,4				
14	34	225	6,6				
15	122	783	6,4				