Daten erheben, bearbeiten und auswerten - Erfahrungen mit einem Reaktionstest bei Schülern und Studenten

Von HEINZ TRAUERSTEIN, Bielefeld

Zusammenfassung: Im 1. Teil des Aufsatzes werden drei Unterrichtsstunden in einem 8. Schuljahr beschrieben. Ein motivierendes Zufallsexperiment wird im Unterricht geplant und durchgeführt. Im Anschluß daran werden die gewonnenen Daten geordnet, verschiedene Kennwerte erarbeitet und ein Vergleich zwischen den Testleistungen der Jungen und Mädchen durchgeführt.

Im 2. Teil des Aufsatzes berichtet der Verfasser von seinen Erfahrungen mit dem gleichen Zufallsexperiment in drei verschiedenen Seminaren mit Lehramtsstudenten. Neben weiteren Techniken zur Beschreibung der Daten (5-Zahlenzusammenfassung und Kastenschaubild) werden verschiedene Hypothesen aufgestellt, die mit Hilfe der Binomialverteilung zufallskritisch geprüft werden. Es wird gezeigt, wie man aus einem unübersichtlichen und stark streuenden Datensatz statistisch signifikante Regelmäßigkeiten herausfiltern kann.

1 Einleitung

Schon in der Grundschule sollen die Kinder erste Erfahrungen sammeln, wie man Daten erhebt, in Tabellen anordnet und in Stabdiagrammen veranschaulicht. Diese Erfahrungen werden in der Sekundarstufe I fortgesetzt. In den Richtlinien für Hauptschulen in NRW wird der Bereich Stochastik zwar nur als zusätzliches Angebot (Additum) für die Klassen 5 und 6 aufgeführt, elementare stochastische Sachverhalte sollen aber im Rahmen der Anwendungsorientierung auch in den anderen Bereichen des Mathematikunterrichts behandelt werden. In Klasse 10 werden dann stochastische Inhalte für alle Schüler und Schülerinnen verpflichtend in einem eigenen Themenkreis zusammengefaßt und weiterentwickelt. Im Bereich der Statistik [1] sollen un-

[1] Für die Klassen 10 vom Typ B ist neben der Statistik noch ein Kapitel über Wahrscheinlichkeitsrechnung vorgesehen

Stochastik in der Schule, **15**(1995), Nr. 1, S. 13-34

ter anderem die folgenden Inhalte behandelt werden: „Statistische Erhebungen durchführen und auswerten; Daten in Tabellen und Schaubildern darstellen. Verschiedene Kennwerte (z.B. Modalwert, Zentralwert, arithmetisches Mittel, ...) einer Verteilung bestimmen; das arithmetische Mittel als zu groben Kennwert erfahren; die Bedeutung einzelner Streumaße (z.B. Spannweite) für die Interpretation einer Verteilung kennen." (Kultusministerium NRW 1989, S. 68)

Wichtig für die Behandlung dieser Inhalte ist es, daß die Schüler und Schülerinnen die im Unterricht benutzten Daten für sich selbst als bedeutungsvoll ansehen. Daher sollte nicht mit fiktiven Zahlen gearbeitet werden, sondern mit realen, nach Möglichkeit auch selbst erhobenen Daten. Nach meinen Erfahrungen ist es besonders motivierend, wenn die Werte etwas über den einzelnen Schüler bzw. die einzelne Schülerin aussagen. Jeder kann dann den Verbleib „seiner Zahlen" in Tabellen, Diagrammen und Kennwerten mitverfolgen. Geeignete Themenbereiche für solche Fragestellungen sind z.B.: Taschengeld, Fernsehgewohnheiten, Schulweg, sportliche Leistungen, Körpergröße und Gewicht. (Vgl. Borovcnik 1987; DIFF-Arbeitsgruppe 1980, 81, 82 und 83; Landesinstitut 1986)

Den Vorschlag für den nachfolgend dargestellten „Reaktionstest" habe ich dem Buch von Borovcnik/Ossimitz (1987, S. 79ff) entnommen und in einem 8. Schuljahr an einer Realschule durchgeführt. Das gleiche Experiment erwies sich auch in drei verschiedenen Veranstaltungen zur Lehrerausbildung als außerordentlich motivierend. An den gewonnen Daten konnten dann auch Fragestellungen zum Testen von Hypothesen demonstriert werden.

2 Planung, Durchführung und Auswertung des Reaktionstests in einem 8. Schuljahr

Im folgenden werden die ersten drei Stunden einer acht Stunden umfassenden Unterrichtsreihe zur Einführung in die Statistik geschildert. Die 13 Jungen und 14 Mädchen der Klasse 8c der Realschule Brackwede hatten bisher noch keine Vorkenntnisse zu diesem Thema.

2.1 Die erste Unterrichtsstunde

In dieser Stunde sollte das vorhergesehene Experiment geplant und durchgeführt werden. Jeder Teilnehmer sollte abschließend einen ersten Wert für seine persönliche Reaktionszeit ermitteln.

Zu Beginn bat der Lehrer - in dieser Stunde war es ein Student - die Schüler, sich im Halbkreis anzuordnen. Die erste Reihe saß auf ihren Stühlen, die zweite Reihe auf den dahinterstehenden Tischen. Erklärt wurde das Experiment zur Messung der Reaktionsschnelligkeit dadurch, daß der Lehrer den Versuch mehrere Male mit einem Schüler durchführte.

Der Versuchsleiter (VL) hielt ein 30cm-Lineal an der 30 cm Marke fest und ließ es frei nach unten hängen. Die Versuchsperson (VP) hielt Daumen und Zeigefinger in der Höhe der 0 cm-Markierung neben das Lineal. Der VL ließ das Lineal ohne Vorankündigung los. Die VP versuchte, das fallende Lineal möglichst schnell mit den beiden wartenden Fingern zu schnappen. Der Fallweg konnte auf der Zentimeterskala des Lineals abgelesen werden.

Nach dieser Demonstration kam von den Schülern der Einwand, daß man damit ja nicht die Reaktionszeit gemessen habe. Der Lehrer versprach, am Ende der Stunde eine Folie mit einem Funktionsgraphen zu zeigen, wo die Zuordnung zwischen den Fallstrecken und den Reaktionszeiten dargestellt war.

Da alle Schüler diesen Test innerhalb von Dreiergruppen durchführen sollten, wurden sie aufgefordert, genaue Regeln für die Ausführung dieses Experiments festzulegen, um für alle Versuchspersonen (VPn) die gleichen Bedingungen zu schaffen. In einem Unterrichtsgespräch wurden verschiedene Störeinflüsse diskutiert und die folgenden Regelungen vereinbart:

- Daumen und Zeigefinger der VP sollen anfangs jeweils etwa 2 cm vom Lineal entfernt sein.
- Die VP sagt „fertig", wenn sie bereit ist.
- Der VL läßt nach dem „Fertig" der VP eine bis maximal zehn Sekunden verstreichen, bevor er das Lineal losläßt.
- Die Wartezeiten des VL sollen möglichst unregelmäßig variieren.
- Jede VP führt das Experiment 20 mal durch.
- Der VL liest die Fallstrecke oberhalb des Daumens ab und rundet dabei auf den nächsten ganzen Zentimeter.
- Jedes Mitglied einer Dreiergruppe ist einmal VP, einmal VL und einmal Protokollant, der die 20 Meßwerte notiert und die Einhaltung der aufgestellten Regeln überwacht.

- Auch wenn das Lineal von der VP nicht festgehalten werden kann und es zu Boden fällt, ist das ein gültiger Versuch. Im Protokoll wird ein Strich eingetragen.
- Die Hand der VP wird auf dem Tisch oder dem Knie aufgelegt, damit sie während des Fangens nicht nach unten geführt werden kann.

Die letzten beiden Regeln wurden erst nach einigen weiteren Demonstrationen des Versuchs aufgestellt, bei denen der Lehrer VP war und sich nicht diesen Vereinbarungen entsprechend verhalten hatte. Die obigen Festlegungen wurden stichwortartig an der Tafel festgehalten.

Im Anschluß an das Unterrichtsgespräch wurde die in der Klasse übliche Sitzordnung wieder hergestellt, und die Schüler führten in Dreiergruppen das Experiment durch. Nach meinen Beobachtungen waren die Mädchen und Jungen mit Eifer und Ehrgeiz bei der Sache.

Zwei von den Teilnehmern, die zuerst fertig waren, schrieben ihre protokollierten Werte an die Tafel.

Oliver: 8, 21, 10, 12, 2, 16, 1, 19, 27, -, 13, 1, 10, 18, 6, 7, 20, 18, 9, 17

Nicole: 16, 20, 13, 14, 17, 8, 11, 12, 17, 19, 11, 16, 9, 15, 20, 18, 28, 12, 10, 21

Da die Unterrichtsstunde fast vorüber war, fiel die abschließende Diskussion über die Meßreihen kürzer als geplant aus. Man einigte sich darauf, jeder VP, wie bei der Berechnung des Notendurchschnitts, das arithmetische Mittel als bestmöglichen Wert zuzuordnen. Für die Striche, bei denen das Lineal durchgefallen war, wurden 32 cm Fallstrecke eingesetzt. Mit Hilfe des Taschenrechners ergab das für Oliver eine mittlere Fallstrecke von 13,35 cm und für Nicole 15,35 cm. Abschließend wurde die Folie mit dem Funktionsgraphen auf den Tageslichtschreiber gelegt, so daß jeder Schüler zu seinem Mittelwert die dazugehörige Fallzeit des Lineals (Reaktionszeit) ablesen konnte (siehe Abb. 1 nächste Seite).

Natürlich kann man die Fallzeit aus der Formel $s = \frac{g}{2}t^2$ berechnen. Dabei ist s der Fallweg in m, t die Fallzeit in sec und g die Gravitationskonstante 9,81 m/sec². Da die Klasse aber noch keine quadratischen Funktionen behandelt hatte, den Schülern aber Funktionsgraphen von den linearen Funktionen her vertraut waren, wurde dieser Weg gewählt. Die Krümmung der Kurve wurde mit dem Hinweis auf das immer schneller fallende Lineal begründet.

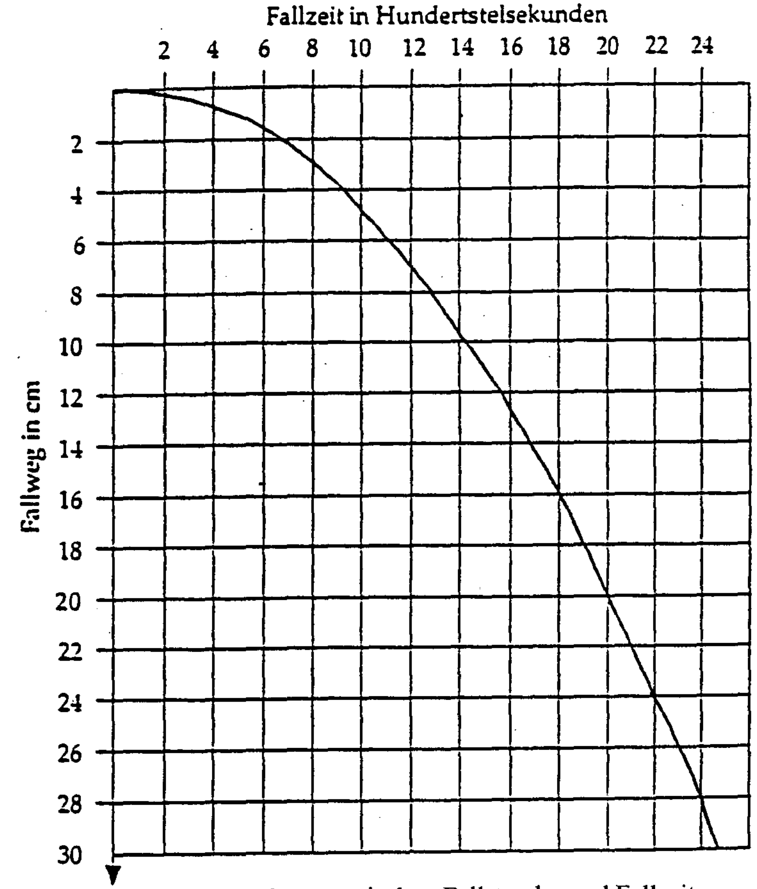


Abb. 1: Zuordnung zwischen Fallstrecke und Fallzeit

2.2 Die zweite Unterrichtsstunde

In dieser Stunde sollten die Schüler das Stengel-und-Blatt-Bild kennenlernen und den Zentralwert einer Datenmenge bestimmen.

Die Studentin, die in dieser Stunde unterrichtete, hatte die Fallzeiten von allen Schülern und Schülerinnen auf einer Folie in Tabellenform zusammengefaßt. Zunächst wurde die Berechnung des arithmetischen Mittels noch einmal wiederholt und das Ablesen der dazugehörenden Fallzeit am Funktionsgraphen deutlich gemacht.

Zur weiteren Auswertung des Reaktionstests wurde dann am Beispiel der Werte von Oliver gezeigt, wie man die Daten noch einfacher und übersichtlicher gleich in der Form eines Stengel- und Blatt-Bildes aufschreiben kann (vgl. Engel 1987, S. 161).

0	2	1	1		
0	8	6	7	9	
1	0	2	3	0	
1	6	9	8	8	7
2	1	0			
2	7				
3	-				

← Tab. 1: Stengel- und Blatt-Bild der Werte von Oliver
Der senkrechte Strich trennt die Zehner- und Einerziffern der einzelnen Meßwerte. Links vom Strich stehen die Zehner, sie bilden den Stengel, und rechts vom Strich werden die Einer notiert, sie bilden die Blätter.

In der ersten Zeile wurden alle Daten von 0 bis 4 einschließlich eingetragen, in diesem Falle also die drei Werte 2, 1 und 1. Die Größen von 5 bis 9 standen in der zweiten Zeile. In der dritten Zeile sind die 4 Werte 10, 12, 13 und 10 festgehalten worden, und in der letzten Zeile wurde durch den Strich kenntlich gemacht, daß das Lineal einmal auf die Erde gefallen ist und die Fallstrecke über 30 cm betrug.

Aus diesem Bild konnten die Schüler den größten und kleinsten Wert, den zweitgrößten und zweitkleinsten Wert usw. ablesen. So gelangten wir zu den beiden Werten in der Mitte, dem 10. Wert von oben und dem 10. Wert von unten. Die Zahl, die genau zwischen diesen beiden mittleren Werten 12 und 13 liegt, bezeichneten wir als Zentralwert Z [2]. Dieser Wert betrug bei Oliver Z = 12,5 cm.

Um Fehler bei der Bestimmung des Zentralwertes zu vermeiden, ordneten wir die Zahlen (Blätter) in den einzelnen Zeilen der Größe nach:

0							
0	8	9					
1	3	4	1	2	1	2	0
1	5	7	7	9	6	5	8
2	0	0	1				
2	8						
3							

Tab. 2: Stengel- und Blatt-Bild der Werte von Nicole

0							
0	8	9					
1	0	1	1	2	2	3	4
1	5	6	6	7	7	8	9
2	0	0	1				
2	8						
3							

Tab. 3: Geordnetes Stengel- und Blatt-Bild der Werte von Nicole

[2] Dieser Ausdruck wird in den meisten Schulbüchern benutzt, während in der Fachliteratur die Bezeichnung Median vorherrscht.

Aus dem geordneten Stengel-und-Blatt-Bild ließ sich der Zentralwert von Nicole mühelos ablesen. Es ist das Mittel der beiden Zahlen 15 und 16, also Z = 15,5 cm. Nach einem weiteren Beispiel, das die Klasse gemeinsam an der Tafel durchgeführt hatte, bestimmte jeder Schüler den Zentralwert seiner Daten mit Hilfe des geordneten Stengel- und Blatt-Bildes. Anschließend wurde das Verfahren auch auf andere Beispiele mit unterschiedlicher Anzahl von Meßwerten ausgedehnt und die folgende Definition an der Tafel festgehalten:

Der *Zentralwert* einer Menge von N Zahlen, die ihrer Größe nach geordnet sind, ist der Wert in der Mitte (bei ungeradem N) oder das arithmetische Mittel der beiden Werte in der Mitte (bei geradem N).

In der abschließenden Diskussion wurden das arithmetische Mittel und der Zentralwert einer Reihe von Meßwerten miteinander verglichen. Den Schülern wurde die Frage gestellt, welcher der beiden Werte besser zur Bestimmung der Reaktionszeit einer VP geeignet sei. Mit Unterstützung durch die Lehrerin wurden folgende Argumente für die Benutzung des Zentralwertes herausgearbeitet:

- Wenn das Lineal nicht gefangen werden kann und zu Boden fällt, braucht man keinen fiktiven Wert anzunehmen, um den Zentralwert zu berechnen.
- Die extremen Werte (Ausreißerwerte) unter 5 cm und über 30 cm sind eher durch Glück bzw. durch Unaufmerksamkeit zu erklären und sollten für die Bestimmung der Reaktionszeit von geringer Bedeutung sein.

Als Hausaufgaben wurden unter anderem die folgenden zwei Aufgaben gestellt, die deutlich machen sollten, daß das arithmetische Mittel von Ausreißerwerten stark beeinflußt wird.

1) Bei den Achtelfinalspielen um den DFB-Pokal werden von den einzelnen Plätzen die folgenden Zuschauerzahlen gemeldet: 8000, 12000, 13000, 15000, 19000, 23000, 55000, 70000. Berechne das arithmetische Mittel und den Zentralwert.
2) Bei der überregionalen Preiserhebung der Stiftung Warentest für eine bestimmte Kleinbildkamera werden die folgenden Preise in DM ermittelt: 382, 378, 298, 308, 379, 399, 389. Als mittlerer Preis wird 379,- DM angegeben. Welchen Wert benutzt die Stiftung Warentest?

2.3 Die dritte Unterrichtsstunde

Thema der dritten Unterrichtsstunde war ein Vergleich zwischen den Leistungen der Jungen und Mädchen beim Reaktionstest.

Zunächst wurde bei der Besprechung der Hausaufgaben deutlich, daß ein einzelner Kennwert wie das arithmetische Mittel oder der Zentralwert nur eine sehr begrenzte und möglicherweise auch verfälschende Information wiedergibt. Man sollte zumindest noch den größten und kleinsten Wert der Verteilung angeben oder am besten gleich das Stengel-und-Blatt-Bild.

Danach wurde die Frage gestellt, ob die Jungen oder die Mädchen beim Reaktionstest besser abgeschnitten hätten. Man einigte sich darauf, für jede VP den Zentralwert der 20 Versuche als geeignete Testleistung beim Reaktionstest zu benutzen und die Werte der Jungen und Mädchen anzuschreiben:

Jungen: 12,5; 18,5; 16; 16; 15; 15; 16; 16; 16,5; 9; 17; 23; 20,5
Mädchen: 15; 12; 15; 18; 11; 15; 16; 13; 17; 10; 13; 15,5; 11,5; 18;

Der Lehrer zeigte an der Tafel, wie man diese Werte in ein zweiseitiges Stengel-und-Blatt-Bild einträgt, um die beiden Gruppen besser vergleichen zu können: siehe Tab. 4 nächste Seite. Der „Stengel" des Bildes besteht jetzt aus den ganzzahligen Anteilen vor dem Komma und die „Blätter" aus der ersten Nachkommastelle der einzelnen Meßwerte [3].

Folgende Kennwerte bei den Verteilungen wurden bestimmt und in einer Tabelle (Tab. 5 nächste Seite) an der Tafel zusammengestellt.

Die Mädchen hatten beim Reaktionstest eindeutig besser abgeschnitten. Einziger Lichtblick für die Jungen: sie stellten den Klassenbesten. Ein Junge vermutete, daß das schlechte arithmetische Mittel der Jungen möglicherweise auf den Ausreißerwert 23,0 zurückzuführen sei. Um das zu überprüfen, wurde dieser Wert bei den Jungen gestrichen und dann das arithmetische Mittel der restlichen 12 Werte berechnet. Es lag mit 15,67 zwar unter dem Wert von 16,23, war aber immer noch schlechter als der entsprechende Wert der Mädchen. Auch wurde deutlich, daß diese Streichung die anderen beiden Mittelwerte, den Zentralwert und den häufigsten Wert, nicht verändert hätte.

[3] Es gibt noch eine Reihe anderer Möglichkeiten, die Konstruktion des Stengel-und-Blatt-Bildes an die jeweiligen Daten anzupassen. (Siehe dazu Biehler 1982 und Borovcnik 1987)

Tab. 4: Zweiseitiges Stengel-und-Blatt-Bild

Jungen		Mädchen
0	9	
	10	0
	11	0 5
5	12	0
	13	0 0
	14	
0 0	15	0 0 0 5
5 0 0 0 0	16	0
0	17	0
5	18	0 0
	19	
5	20	
	21	
	22	
0	23	

Tab. 5: Vergleich der verschiedenen Kennwerte

	Jungen	Mädchen
Anzahl	13	14
Zentralwert	16,0 cm	15,0 cm
arithmetisches Mittel	16,23 cm	14,29 cm
häufigster Wert	16,0 cm	15,0 cm
kleinster Wert	9,0 cm	10,0 cm
größter Wert	23,0 cm	18,0 cm

Beim Vergleich der beiden Gruppen fiel auf, daß die Werte der Jungen in einem breiteren Bereich verteilt lagen als die der Mädchen. Das gab Anlaß für die Definition der Spannweite einer Menge von Zahlen als der Differenz zwischen dem größten und dem kleinsten Wert. Die Spannweite der Jungen betrug 14 cm und die der Mädchen 8 cm. Diese Werte wurden auch noch in der obigen Tabelle eingetragen.

Man hätte im Anschluß daran die Bestimmung der Quartile und die Quartildistanz als weiteres Streuungsmaß einführen können. Wir haben in dieser Stunde zugunsten eines weiteren Übungsbeispiels aus dem Schulbuch (Traeger 1973) darauf verzichtet. Damit war die Behandlung des Reaktionstests in dieser Klasse abgeschlossen. In den nächsten zwei Stunden der Unterrichtseinheit wurde eine weitere Erhebung durchgeführt, bei der die Schüler die Längen verschiedenartiger Gegenstände und Linien schätzen sollten. Am Beispiel dieser Daten wurde neben den bisher bekannten Kennwerten noch die Quartildistanz als weiteres Streuungsmaß behandelt. Die letzten drei Stunden hatten dann statistische Erhebung mit zwei Merkmalen (Körpergröße und Gewicht) zum Thema.

3 Erfahrungen mit dem Reaktionstest bei Studenten

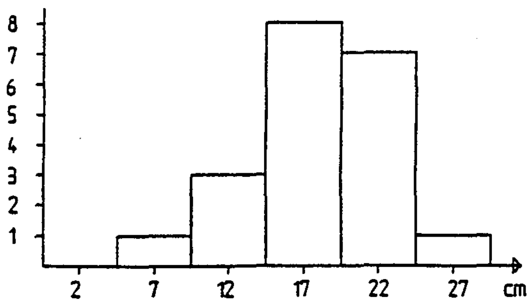
3.1 Die 5-Zahlenzusammenfassung und das Kastenschaubild

Nach den positiven Erfahrungen im 8. Schuljahr habe ich diesen Reaktionstest in drei Seminaren im Rahmen der Lehrerausbildung durchgeführt. Auch die Studentinnen und Studenten für das Lehramt für die Primarstufe haben das Experiment mit Vergnügen durchgeführt. Die in Partnerarbeit erhobenen Meßwerte wurden in eine vorbereitete Tabelle an der Tafel eingetragen und für die nächste Sitzung auf Arbeitsblättern vervielfältigt (s. Anhang). Die Erhebung und Bearbeitung der Daten entsprach in etwas gestraffter Form dem in Abschnitt 2 geschilderten Vorgehen in der Schulklasse. Neben dem Stengel-und-Blatt-Bild sollten die Studenten und Studentinnen jeweils das zugehörige Histogramm zeichnen. A. Engel (1987, S. 161) vermerkt über das Stengel-und-Blatt-Bild sicher zur Recht: „Das entstandene Bild ist eine Verbesserung des traditionellen Histogramms, da es dieselbe Information genauer und mit weniger Arbeit liefert." Beim Anfertigen des Histogramms hatten mehrere Seminarteilnehmer Schwierigkeiten, die Klassenmitteln und Klassengrenzen zu erkennen.

0								
0	6							
1	2	3	4					
1	5	6	6	6	7	8	9	9
2	1	1	2	3	3	3	4	
2	9							
3								

← Tab. 6: Geordnetes Stengel-und-Blatt-Bild der Fallstrecken von VP ①

↓ Abb. 2: Histogramm der Fallstrecken von VP ①



Als Mittelwerte wurden das arithmetische Mittel, der Zentralwert (Median) und der häufigste Wert (Modalwert) anhand der Daten aus dem Reaktionstest eingeführt. Als weitere Kennwerte einer Verteilung kamen noch die Quartile hinzu. Sie wurde wie folgt definiert: Eine Menge von N Zahlen, die ihrer Größe nach geordnet sind, wird durch den Zentralwert in zwei gleiche Hälften geteilt. Bei ungeradem N rechnet man die mittlere Zahl in jeder der beiden Hälften mit. Der Zentralwert der linken Hälfte der Daten heißt 1. Quartil und der Zentralwert der rechten Hälfte ist das 3. Quartil. Durch das 1. und 3. Quartil werden jeweils das erste und das letzte Viertel

der Daten abgetrennt, so daß zwischen dem 1. und dem 3. Quartil etwa die Hälfte der Werte liegen. Die Differenz Q_3 - Q_1 heißt Quartildistanz und wird als Maß für die Streuung der Stichprobe benutzt.

Jede VP trug dann die wichtigsten Kennwerte ihrer 20 Versuche in die sogenannte „Fünfzahlen-Zusammenfassung" ein (vgl. Biehler 1982, S. 44).

N				
T(Z)		Z		
T(Q)	Q_1		Q_3	Quartildistanz
1	Min.		Max.	Spannweite

Abb. 3: Fünfzahlen-Zusammenfassung

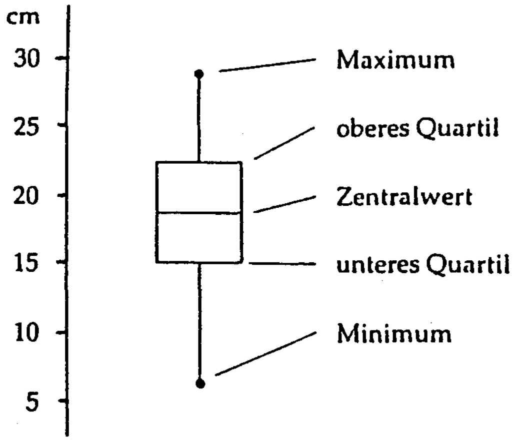
Dabei bedeuten T(Z) bzw. T(Q) die Tiefe von Z bzw. Q. Diese Zahlen geben an, inwieweit Z bzw. Q im „Inneren" des Datensatzes liegen. Für die 20 Zahlen der VP ① ist T(Z) = 10 h (h benutzt man als Abkürzung für ½). Das bedeutet, daß Z zwischen dem 10. und dem 11. Wert in der geordneten Datenreihe liegt. T(Q) = 5 h, da das 1. Quartil zwischen dem 5. und 6. Wert von links und das 3. Quartil zwischen dem 5. und 6. Wert von rechts zu finden ist. Für die Tiefen [4] gilt $T(Z) = \frac{N+1}{2}$ und $T(Q) = \frac{T[(Z)]+1}{2}$. Die eckige Klammer (Gaußklammer) um T(Z) gibt an, daß hier nur der ganzzahlige Teil von T(Z) berücksichtigt wird.

20				
10 h		18,5		
5 h	15,5		22,5	7
1	6		29	23

Abb. 4: VP ① Fünfzahlen-Zusammenfassung (Fallstrecken in cm)

Zur weiteren Veranschaulichung der Fünfzahlen-Zusammenfassung haben wir das Kastenschaubild gezeichnet (vgl. Biehler 1982, S. 50ff und Borovcnik 1987, S. 41f). Besonders beim Vergleich mehrerer Kastenschaubilder waren Unterschiede und Gemeinsamkeiten der entsprechenden Verteilungen gut zu erkennen.

[4] Die Werte für die Tiefe von Z und Q werden in Borovcnik (1987, S. 37 f) etwas anders angegeben, so daß z.B. bei N = 17 oder N = 19 andere Quartile möglich sind. Ich habe mich hier an die bei Biehler (1982) beschriebene Vorgehensweise von Tukey angelehnt.



← Abb. 5: Kastenschaubild VP ①

3.2 Leistungsvergleich zwischen Schülern und Studenten im Reaktionstest

Bei der Auswertung des Reaktionstests in den Seminaren stellte sich natürlich die Frage, wie die eigene Gruppe im Vergleich mit den Schülern und den anderen Seminargruppen abgeschnitten hatte. Daher wurde jeweils die Fünfzahlen-Zusammenfassung aufgestellt. Als Testleistung jeder VP wurde wieder der Zentralwert ihrer Fallstrecken verwendet.

Klasse 8c

27			
14	15,5		
7h	13,0	16,5	3,5
1	9,0	23,0	4,0

Seminar: Stochastik WS 1991/92

17			
9	15,5		
5	14,5	17,5	3,0
1	11,5	19,5	8,0

Seminar Stochastik SS 1993

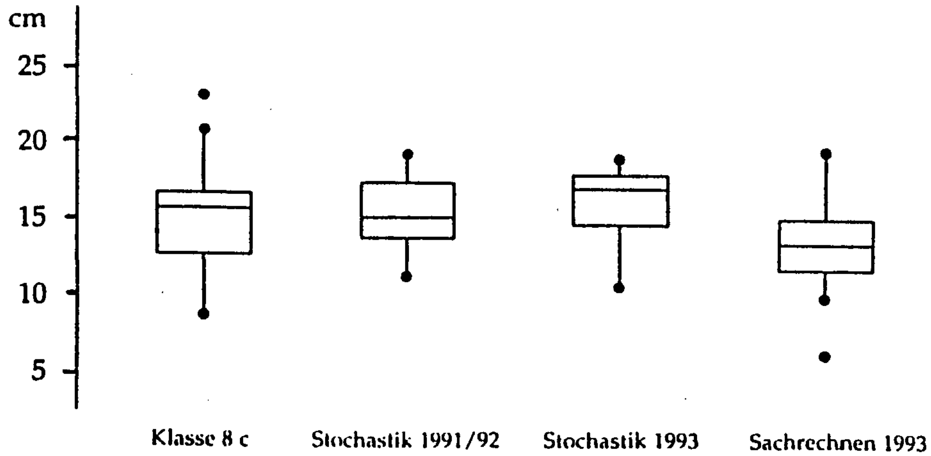
22			
11h	17,0		
6	15,0	18,0	3,0
1	11,0	19,0	8,0

Seminar: Sachrechnen SS 1993

18			
9h	13,75		
5	11,5	16,0	4,5
1	6,5	18,5	12,0

Abb. 6: Fünfzahlenzusammenfassungen für die verschiedenen Gruppen

Abb. 7: Kastenschaubilder der 4 Gruppen (Zentralwerte der Fallstrecken in cm)



Die Kastenschaubilder zeigen anschaulich die Mittelwerts- und Streuungsunterschiede zwischen den einzelnen Gruppen. Auch die unterschiedliche Lage des Zentralwertes im mittleren 50 %-Intervall wird deutlich. Die beste Reaktionsleistung von allen Teilnehmern mit einem Zentralwert von 6,5 cm bei 20 Versuchen erzielte ein Student der 4. Gruppe. Da die nächstbeste Leistung erst bei 9,5 cm lag, wurde die senkrechte Linie im Kastenschaubild nicht bis zu dem Ausreißerwert von 6,5 cm durchgezeichnet. Ähnlich wurde bei der schlechtesten Testleistung von 23,0 cm verfahren [5].

3.3 Läßt sich ein Trainings- oder Ermüdungseffekt nachweisen?

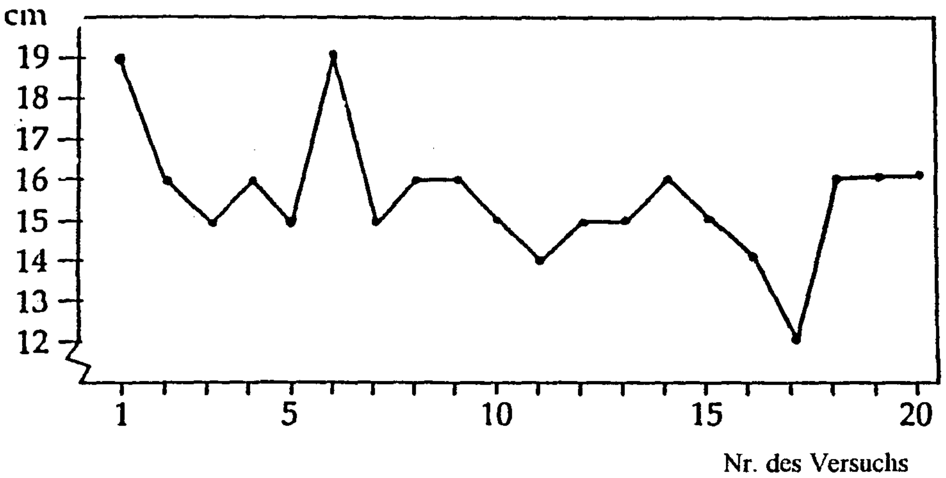
Jede Versuchsperson hat 20 mal nacheinander versucht, das herunterfallende Lineal möglichst schnell aufzufangen. Es ist gut nachvollziehbar, wenn im Laufe der vielen Versuche die Konzentration einer VP nachläßt und ein „Ermüdungseffekt" eintritt. Andererseits kann man sich auch vorstellen, daß in einer Aufwärmphase zunächst die entsprechenden Körperfunktionen

[5] Die auslaufenden Striche oberhalb und unterhalb des Kastens sollen den Bereich kennzeichnen, in dem die Verteilung einen „geschlossenen" Eindruck macht. Die Grenzen werden per Augenmaß aus dem Stengel-und-Blatt-Bild abgelesen und bleiben der subjektiven Einschätzung überlassen. Die Ausreißerwerte werden als einzelne Punkte eingezeichnet (vgl. Borovcnik 1987, S. 41).

aktiviert werden müssen, man sich allmählich besser auf diese Aufgabe einstellt und die Testleistung ansteigt. Wir haben dies als „Trainingseffekt" bezeichnet.

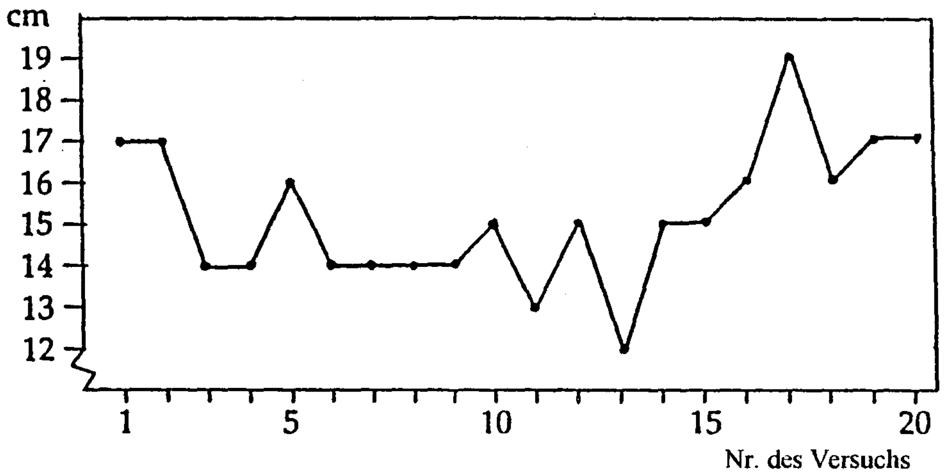
Wenn ich die Seminarteilnehmer aufforderte, ihre eigenen Erfahrung mit dem Test zu reflektieren und Vermutungen über einen möglichen Trainings- oder Ermüdungseffekt zu äußern, stieß ich auf merkliche Zurückhaltung. Die eigenen Testergebnisse zeigten starke Zufallsschwankungen und ließen keinerlei Regelmäßigkeit erkennen. Auch ein Blick auf die Ergebnistabelle der gesamten Gruppe bestärkte die meisten Teilnehmer in der Annahme, daß aus diesen Zahlen keine solche Gesetzmäßigkeit abzuleiten war. Um einen möglicherweise vorhandenen Trainings- oder Ermüdungseffekt deutlich werden zu lassen, ermitteln wir im WS 1991/92 für jeden einzelnen Versuch den Zentralwert aller 17 Versuchspersonen, um so einen Eindruck von der Leistung der gesamten Gruppe im Verlauf der 20 Versuche zu erhalten. Das Ergebnis wurde in einem Polygonzug dargestellt.

Abb. 8: Mediane der Fallstrecken der einzelnen Versuche im WS 1991/92



Zum Vergleich stellte ich den entsprechenden Polygonzug für die Klasse 8c zusammen.

Abb. 9: Mediane der Fallstrecken der einzelnen Versuche in der Klasse 8c



Beide Kurven wiesen starke zufallsbedingte Schwankungen auf, so daß kein Trainings- oder Ermüdungseffekt erkennbar war. Auch eine mögliche Glättung der Kurven (vgl. Spiegel, S. 285) schien mir nicht lohnenswert.

In dem Stochastikseminar im SS 1993 äußerte eine Studentin die Vermutung, daß sie sich trotz aller Leistungsschwankungen in der ersten Hälfte der Versuche gesteigert habe und dem Ende hin ihre Leistungen aber wieder schlechter geworden seien. Dieser Beobachtung schlossen sich einige Teilnehmer an, während die Mehrzahl skeptisch blieb. Um der Fragestellung nachzugehen, habe ich im SS 1993 einen anderen Weg eingeschlagen als im WS 1991/92.

Die starken zufälligen Schwankungen zwischen den einzelnen Versuchen sollten nicht so durchschlagen. Deshalb teilten wir die 20 Versuche in 5 Vierergruppen ein. Den Zentralwert der ersten 4 Versuche einer VP wählten wir als Maß für ihre „Anfangsleistung". Aus den mittleren 4 Versuchen (Nr. 9, 10, 11 und 12) ergab sich die „Halbzeitleistung", und den Median der letzten vier Versuche bezeichneten wir als „Endleistung" einer VP.

Wir präzisieren nun unsere Vermutung:

In der ersten Testhälfte tritt bei der Mehrzahl der Teilnehmer ein Trainingseffekt ein, und in der zweiten Testhälfte macht sich ein Ermüdungseffekt bemerkbar.

Da in dem Seminar die Binomialverteilung und erste Beispiele zum Testen von Hypothesen (mit Hilfe der Binomialverteilung) bereits behandelt worden waren, leiteten wir die folgenden statistisch überprüfbaren Hypothesen ab:

<u>Nullhypothese H</u>: Die Wahrscheinlichkeit p, daß eine VP ihre „Anfangsleistung“ zur „Halbzeit“ verbessert, ist $p = \frac{1}{2}$.

<u>Alternativhypothese A</u>: $p > \frac{1}{2}$ oder $p < \frac{1}{2}$.

Wir wählten hier den zweiseitigen Test, da auch eine mehrheitliche Verschlechterung zur Halbzeit nicht ausgeschlossen werden konnte.

Nachfolgend sich die „Anfangsleistung“, die „Halbzeitleistung“ und die „Endleistung“ jedes Teilnehmers am Stochastikseminar im SS 1993 zusammengefaßt.

Nummer der VP	1	2	3	4	5	6	7	8	9	10
Anfangsleistung	17,5	16,5	16	21	21,5	18,5	22,5	23	19,5	20,5
Halbzeitleistung	19	16,5	19,5	11	15	8,5	16	14,5	16	15
Endleistung	21	16,5	19	16,5	17,5	9	16,5	21,5	15	18,5

Nummer der VP	11	12	13	14	15	16	17	18	19	20	21	22
Anfangsleistung	19,5	18,5	15,5	11,5	12,5	20,5	18	20,5	21	21	18,5	15,5
Halbzeitleistung	12,5	15	13,5	11	10	14,5	16,5	14,5	19	12	22	15,5
Endleistung	12,5	15,5	16	17,5	15	14,5	16,5	15	17	12,5	17	13,5

Tab. 7: Leistungsverlauf bei den einzelnen Teilnehmern

Von den 22 Teilnehmern hatten X = 17 zur Halbzeit bessere Lösungen gezeigt als in den ersten 4 Versuchen und 3 VPn hatten sich verschlechtert. Bei VP 2 und VP 22 stimmten beide Werte überein. Solche Fälle werden bei diesem Test ausgeschlossen und die Versuchspersonenzahl wird entsprechend reduziert (vgl. Siegel 1956, S. 71).

Um zwischen den beiden Hypothesen H und A zu entscheiden, berechneten wir die Wahrscheinlichkeit α dafür, daß solch ein extremes oder noch extremeres Ergebnis unter der Annahme von H eintritt.

Da die Anzahl der Verbesserungen binomial verteilt ist, mit N = 20 und $p = \frac{1}{2}$, und wir den zweiseitigen Test gewählt hatten, gilt:

$$\alpha = 2 \cdot w(X \geq 17) = 2 \cdot \sum_{k=17}^{20} \binom{20}{k} \cdot (\tfrac{1}{2})^{20} = 2 \cdot \frac{1351}{1048576} = 0{,}0026 .$$

Dieses Ergebnis ist signifikant auf dem 1 %-Niveau. Wir haben daher H verworfen und die Alternativhypothese angenommen. Die Wahrscheinlichkeit dafür, daß die Halbzeitleistung einer VP besser ist als ihre Anfangsleistung, ist größer als $\frac{1}{2}$. Bei der Mehrzahl der Versuchspersonen ist also in der ersten Testhälfte eine Verbesserung zu erwarten. Das heißt, daß ein Trainingseffekt nachgewiesen wurde.

Es stellte sich für mich die Frage, ob dieser Effekt in den anderen Gruppen auch bestätigt werden konnte. Ich faßte die drei übrigen Gruppen zusammen und zählte 41 Verbesserungen und 19 Verschlechterungen. Bei jetzt einseitiger Fragestellung erhielt ich:

$$\alpha = 2 \cdot w(X \geq 41) = w\left[z \geq \frac{40{,}5 - 30}{\sqrt{60 \cdot 0{,}5 \cdot 0{,}5}}\right] = w(z \geq 2{,}71) = 0{,}0034 .$$

Für die Berechnung von α habe ich hier mit der Näherung von Laplace-Moivre mit Stetigkeitskorrektur gearbeitet (vgl. Siegel 1956, S. 72 oder DIFF 1981, 1981, S. 33 ff). Da dieses Verfahren mit den Primarstufenstudenten nicht behandelt wurde, habe ich den Wert nur als zusätzliche Bestätigung für den gefundenen Trainingseffekt mitgeteilt.

Als nächstes wurde überprüft, ob in der zweiten Testhälfte ein Ermüdungseffekt nachgewiesen werden kann. Wir formulieren die <u>Nullhypothese H</u>: Die Wahrscheinlichkeit p, daß sich eine VP bei der Endleistung im Vergleich zur Halbzeitleistung verschlechtert, ist $p = \frac{1}{2}$. Die <u>Alternativhypothese A</u> lautet dann: $p \neq \frac{1}{2}$.

Bei 13 VPn trat die erwartete Verschlechterung ein, wie man der obigen Tabelle entnehmen kann, und 5 Teilnehmer verbesserten sich. Die 4 VPn mit unveränderten Werten blieben auch hier unberücksichtigt. Wir testeten wieder zweiseitig und erhielten:

$$\alpha = 2 \cdot w(X \geq 13) = 2 \cdot \sum_{i=13}^{18} \binom{18}{i} \cdot (\tfrac{1}{2})^{18} = 0{,}096.$$

Dieses Ergebnis war also nicht signifikant auf dem 5 %-Niveau. Wir mußten daher vorerst, aus Mangel an Beweisen, H beibehalten.

Will man trotzdem versuchen, diese Hypothese zufallskritisch abzusichern, kann man entweder die Versuchspersonenzahl erhöhen oder versuchen, ein wirkungsvolleres Testverfahren einzusetzen [6]. Da uns die Testergebnisse von den anderen Gruppen zur Verfügung standen, entschieden wir uns für die erste Möglichkeit. Bei 8 von den insgesamt 84 VPn stimmten Halbzeitleistung und Endleistung überein.

Von den verbleibenden 76 VPn zeigten 48 die erwartete Verschlechterung.

$$\alpha = 2 \cdot w(X \geq 48) = 2 \cdot w\left(z \geq \frac{47{,}5 - 38}{\sqrt{19}}\right) = 2 \cdot w(z \geq 2{,}17) = 2 \cdot 0{,}015 = 0{,}03.$$

Dieses Ergebnis ist mit $\alpha = 3$ % auf dem 5 %-Niveau signifikant. Allerdings ist der beobachtete Ermüdungseffekt bei weitem nicht so häufig anzutreffen wie der zuerst nachgewiesene Trainingseffekt [7].

Die Reaktionsleistungen der VPn im Verlauf der 20 Versuche wollten wir noch detaillierter untersuchen [8]. Aus unserer anfangs aufgestellten Vermutung leiteten wir die Hypothese ab, daß das folgende Leistungsverhalten überzufällig oft anzutreffen ist: am Anfang die schlechtesten Reaktionswerte, bei Halbzeit die Bestleistung und am Ende wieder ein Wert, der zwischen der Anfangs- und Halbzeitleistung liegt.

Diesen Leistungsverlauf charakterisieren wir durch die Buchstabenfolge SBM (zuerst schlechteste, dann beste und zuletzt mittlere Leistung). Außer dieser Kombination sind noch die Folgen BMS, BSM, MBS, MSB und SMB möglich. Die Nullhypothese basiert auf der Annahme, daß die Leistungsschwankungen nur zufallsbedingt sind. Das bedeutet für die sechs möglichen Kombinationen, daß sie alle gleichwahrscheinlich sind.

---

[6] Da der hier benutzte Vorzeichentest nicht sehr effektiv ist (vgl. Siegel 1956, S. 75), bieten sich zur weiteren Überprüfung der Hypothese der Vorzeichen-Rangtest von Wilcoxon und als wirkungsvollster Test der Permutationstest für verbundene Paare an (vgl. Engel 1987, S. 75 ff oder Siegel 1956, S. 166 ff).

[7] Von den 80 VPn mit unterschiedlicher Anfangs- und Halbzeitleistung zeigten 58 eine Verbesserung, während sich nur 22 zur Halbzeit verschlechterten.

[8] Zunächst hätte man auch überprüfen können, ob die Bestleistungen besonders oft zur Halbzeit erzielt werden. Für die Teilnehmer aller vier untersuchten Gruppen zusammen hätte sich bei $N = 75$, $p = \frac{1}{3}$ und $X = 41$ eine Irrtumswahrscheinlichkeit von $\alpha < 0{,}0003$ ergeben. Auch dieses Ergebnis wäre hochsignifikant gewesen.

Wir formulierten daher als Nullhypothese H: Die Wahrscheinlichkeit p, daß eine VP den Leistungsverlauf SBM zeigt, ist $p = \frac{1}{6}$.
Die Alternativhypothese A lautet dann $p > \frac{1}{6}$ [9].

Von den 22 Seminarteilnehmern im SS 1993 wurden zunächst 5 ausgeschieden, weil mindestens zwei ihrer drei Werte übereinstimmten und sie deshalb keinem der sechs möglichen Verlaufsmuster zugeordnet werden konnten. Von den verbleibenden 17 VPn zeigten $X = 9$ Teilnehmer die Wertefolge SBM. Wir berechneten die Wahrscheinlichkeit α dafür, daß solch ein extremes oder ein noch extremeres Ergebnis unter der Ausnahme H zustanden kommt. Da X binomialverteilt ist, mit $N = 17$ und $p = \frac{1}{6}$, erhält man:

$$\alpha = w(X \geq 9) = \sum_{k=9}^{17} \binom{17}{k} \left(\frac{1}{6}\right)^k \cdot \left(\frac{5}{6}\right)^{17-k} = 0{,}0007.$$

Wir konnten also die Nullhypothese verwerfen und die Alternativhypothese annehmen.

Es stellte sich wieder die Frage, ob dieses sehr signifikante Ergebnis in den anderen 3 Gruppen bestätigt werden konnte. Faßt man die übrigen Teilnehmer zusammen, so kann man bei 16 von 53 VPn den Verlauf SBM beobachten. Für die entsprechende Irrtumswahrscheinlichkeit gilt dann:

$$\alpha = w\left(z \geq \frac{15{,}5 - 8{,}83}{\sqrt{53 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right) = w(z \geq 2{,}64) = 0{,}0041 \text{ [10]}.$$

---

[9] Man hätte hier auch die beiden Hypothesen etwas vorsichtiger formulieren können: H: Alle sechs möglichen Kombinationen haben jeweils die Wahrscheinlichkeit $p = \frac{1}{6}$. A: Für mindestens eine Kombination gilt $p > \frac{1}{6}$. Dann müßte man das berechnete $\alpha = 0{,}0007$ mit 6 multiplizieren und erhielte immer noch einen Wert, der deutlich unter dem 1 %-Niveau liegt.

[10] Bei größeren Versuchspersonenzahlen - in unserem Fall bei $N \geq 30$ - kann hier auch der Chi-Quadrat-Test benutzt werden, um zu prüfen, ob die beobachteten Häufigkeiten der einzelnen Verlaufskombinationen mit den unter der Nullhypothese erwarteten Häufigkeiten übereinstimmen. Ich habe das für alle 70 VPn, die einem der sechs möglichen Muster zugeordnet werden konnten, durchgeführt.

	BMS	BSM	MBS	MSB	SBM	SMB
beobachtete Häufigkeit	6	5	14	9	25	11

b.w.

Nachdem wir unsere Vermutung über den Leistungsverlauf bei den 20 Versuchen durch die Signifikanztests eindrucksvoll bestätigt fanden, wußten wir, daß der Trainingseffekt in den ersten, der Ermüdungseffekt in der zweiten Hälfte und das Verlaufsmuster BSM überzufällig oft angetroffen werden. Über die Größe dieser Effekte geben die Signifikanztests keinen Aufschluß. Man erhält höchstens indirekt aus der Größe der verschiedenen Irrtumswahrscheinlichkeiten gewisse Anhaltspunkte.

Deshalb wollten wir abschließend die Leistungsunterschiede in den fünf verschiedenen Testphasen in einem Polygonzug veranschaulichen. Für jede Phase werden die Mediane der Leistungen von den Teilnehmern der einzelnen Gruppen eingetragen (siehe Abb. 10 nächste Seite)

Diese Kurven geben trotz der Unterschiede zwischen den einzelnen Gruppen jeweils Hinweise auf das Ausmaß des Trainingseffekts in der ersten Testhälfte und die Größe des Ermüdungseffekts in der zweiten Hälfte. Faßt man die vier Gruppen zusammen, so konnten die Teilnehmer ihre Anfangsleistung zur Halbzeit im Mittel um 2,5 cm (Median der Verbesserungen) steigern und der Abfall in der zweiten Testhälfte betrug im Mittel 1 cm. Die Schwankungen zwischen den Werten der einzelnen VPn sind jedoch beträchtlich.

Damit war die Auswertung des Reaktionstests im SS 1993 abgeschlossen. Für die Teilnehmer und auch für mich selber war dies ein eindrucksvolles Beispiel dafür, wie man aus einem unübersichtlichen und streuenden Datensatz zufallskritisch abgesicherte Regelmäßigkeiten herausfiltern kann.

4 Literatur

BERG, D. U.A.: Statistik und Wahrscheinlichkeitsrechnung G., Mathematik S-2, Freiburg 1980

BIEHLER, R.: Explorative Datenanalyse - Eine Untersuchung aus der Perspektive einer descriptiv-empirischen Wissenschaftstheorie. IDM Materialien und Studien Band 24, Bielefeld 1982

BLANKENAGEL, J. U.A.: Grundkurs Stochastik, Mathematisches Unterrichtswerk für die Sekundarstufe II, Stuttgart 1982

Man errechnet (vgl. Siegel 1956, S. 42 ff oder Siegel 1976, S. 210 ff):

$\chi^2_{emp.} = 22{,}34$ und $\chi^2_{theor.}(df = 5;\ p = 0{,}005) = 16{,}7$.

Auch hier ist das Ergebnis auf dem 1 %-Niveau signifikant.

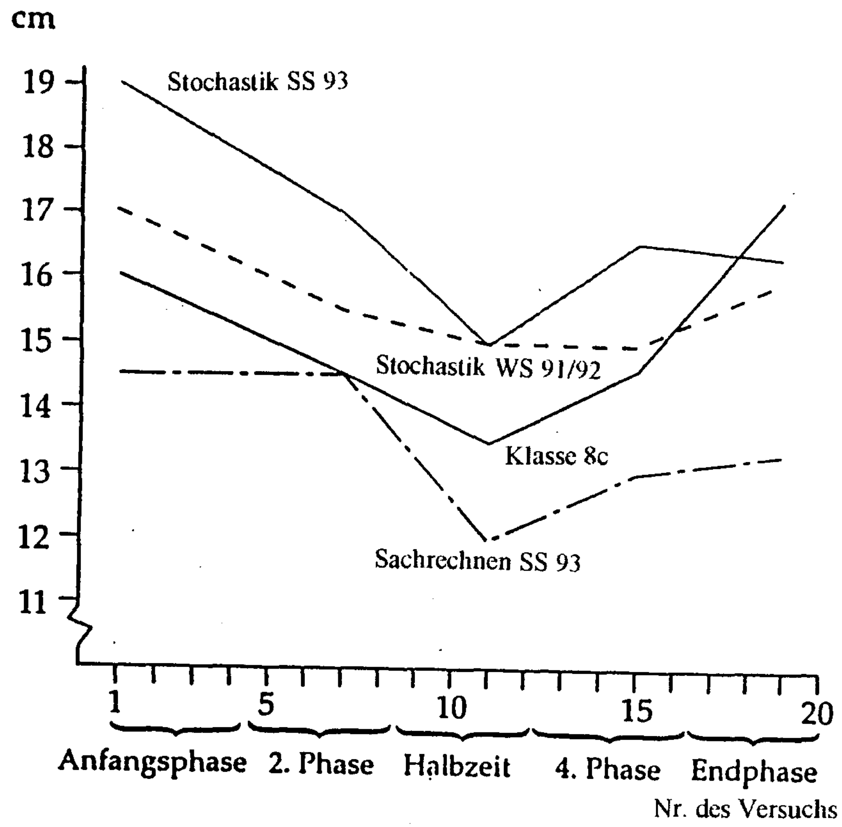


Abb. 10: Leistungen der Gruppen in den 5 verschiedenen Testphasen

BOROVCNIK, M. / OSSIMITZ G.: Materialien zur Beschreibenden Statistik und Explorative Datenanalyse. Wien und Stuttgart 1987

DIEPGEN, K. / KUYPERS, W. / RÜDIGER, K.: Stochastik-Grundkurs. Bielefeld 1989

DIFF-ARBEITSGRUPPE: Beschreibende Statistik. Mathematik, Kurs für Lehrer Sekundarstufe I/Hauptschule, HE 11, Tübingen 1980

DIFF-ARBEITSGRUPPE: Mathematik - Studienbriefe zur Fachdidaktik für Lehrer der Sekundarstufe II, Stochastik MS 1 und MS 4, Tübingen 1980 und 1981

DIFF-ARBEITSGRUPPE: Mathematik, Wahrscheinlichkeitsrechnung und Statistik unter Einbeziehung von elektronischen Rechnern, SR 1 und SR 4, Tübingen 1982 und 1983

ENGEL, A.: Stochastik, Stuttgart 1987

KULTUSMINISTERIUM NRW (HRSG.): Richtlinien und Lehrpläne für die Hauptschule in Nordrhein-Westfalen, Mathematik, Frechen 1989

KULTUSMINISTERIUM NRW (HRSG.): Richtlinien und Lehrpläne für das Gymnasium - Sekundarstufe I - in Nordrhein-Westfalen, Frechen 1993

LANDESINSTITUT FÜR SCHULE UND WEITERBILDUNG (SOEST) (HRSG.): Stochastik in der Klassenstufe 7/8 - Einführung in die Elemente der beschreibenden Statistik, Soest 1986

SIEGEL, S.: Nonparametric Statistics for the Behavioral Sciences, New York u.a. 1956

SPIEGEL, M.: Statistik, Düsseldorf/New York u.a. 1976

TRAEGER, W. U.A.: Mathematisches Arbeitsbuch für das 8. Schuljahr, Frankfurt am Main 1973

5 Anhang: Abb. 11: Fallstrecken beim Reaktionstest im WS 1991/92

VPn → / Versuch ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	19	-	-	-	18	17	13	13	-	21	20	11	25	20	14	18	10
2	12	17	-	15	16	15	15	18	18	22	14	12	20	9	19	25	14
3	23	20	18	17	15	12	14	18	26	14	15	12	20	15	22	11	13
4	22	18	18	14	-	16	16	15	19	12	10	16	18	6	18	22	13
5	24	20	15	15	9	17	19	13	7	9	13	15	16	18	20	21	14
6	23	16	22	19	21	16	14	-	28	9	16	10	20	13	19	22	11
7	16	13	15	10	10	8	13	10	16	7	19	6	16	17	30	18	17
8	21	20	20	11	27	10	9	11	18	9	15	16	17	21	15	20	12
9	16	16	21	12	17	16	8	8	16	11	13	15	16	19	9	24	14
10	19	23	16	11	21	6	3	8	8	20	13	18	15	14	20	18	14
11	6	15	14	14	19	9	17	21	23	10	12	13	22	16	13	18	10
12	14	15	15	17	10	13	11	19	14	9	22	18	20	16	2	23	13
13	13	29	12	17	6	17	12	25	15	8	15	10	18	20	0	15	-
14	15	18	18	10	16	16	13	22	13	14	14	15	17	16	7	20	16
15	29	17	8	15	17	12	15	14	2	11	17	15	22	14	0	15	15
16	21	18	21	11	20	18	15	11	14	12	13	17	17	13	4	9	1
17	23	20	8	7	11	10	12	17	23	12	11	17	19	10	10	20	10
18	17	21	15	11	16	16	15	8	22	-	10	17	16	14	5	21	12
19	18	15	21	14	16	20	16	13	2	2	-	19	19	16	0	16	13
20	16	15	17	13	20	17	20	16	10	14	16	17	20	9	10	19	14