Das »Eins durch Wurzel aus n Gesetz«[1]
Einführung in statistisches Denken auf der Sekundarstufe I

Schon 1972 hat Freudenthal darauf hingewiesen, daß es aus vielen Gründen sinnvoller sei, die Größe von Zufallsschwankungen in endlichen Versuchsserien zu untersuchen, als die Stabilisierung der relativen Häufigkeiten bei »unendlich langen« Serien in den Mittelpunkt des Unterrichts zu stellen. Im folgenden wird Freudenthals Ansatz vertieft und in eine vielfach erprobte Lernsequenz übersetzt. Nur am Rande sei angemerkt, daß C. F. Gauss der Methode der kleinsten Quadrate, deren Nähe zu folgenden Ausführungen auf der Hand liegt, seinen frühen Rum verdankt, weil er mit ihrer Hilfe die Wiederkehr des Planetoiden Ceres vorhersagte, der im Januar 1801 von G. Piazzi kurzzeitig beobachtet worden war.

Verfasser: Dr. Wolfgang Riemer, August-Bebel-Straße 80, 50259 Pulheim

1 Einführung

Die folgende Unterrichtssequenz versucht, an intuitive Vorerfahrungen anknüpfend, Schülerinnen und Schülern der Klassenstufe 9/10 in einige Grundgedanken beurteilender Statistik einzuführen. Wir untersuchen experimentell die Größe von Zufallsschwankungen und studieren quantitativ, wie diese Schwankungen mit größer werdendem Versuchsumfang abnehmen. Die »griffige« Gesetzmäßigkeit: »Die Schwankungsbreite ist umgekehrt proportional zur Wurzel aus dem Versuchsumfang« ist ein Anwendungsbeispiel für Potenzfunktionen und stellt eine Brücke zur Algebra der Jahrgangsstufe 10 her. Verbindet man andererseits dieses Gesetz mit sensorischen Tests (schmeckst Du, siehst Du, hörst Du einen Qualitätsunterschied zwischen . . .?), so wird sich kaum eine Lerngruppe der Faszination entziehen, die mit Entscheidungen unter Unsicherheit und den Grundgedanken beurteilender Statistik verbunden sein kann.

Durch das experimentelle Vorgehen (wir betrachten das $1/\sqrt{n}$-Gesetz als Naturgesetz) wird eine Grundlage für theoretische Untersuchungen in der Sekundarstufe II geschaffen (zentraler Grenzwertsatz). Andererseits ist die Lernsequenz auf einem elementaren Niveau in sich so abgeschlossen, daß auch Schüler, die das Gymnasium nach der Klasse 10 verlassen, einen Einblick in dieses anwendungsnahe Gebiet der Mathematik erhalten.

[1] Vortrag auf der 83. Hauptversammlung des Deutschen Vereins zur Förderung des mathematischen und naturwissenschaftlichen Unterrichts in Bielefeld 1992.

Da die Zeit in der Jahrgangsstufe 10 oft knapp bemessen ist, kann man sich auf den (alles Wesentliche enthaltenden) Fall $p = 0{,}5$ beschränken. Dann läßt sich die Lernsequenz in ca. 5 Wochen (eine Klassenarbeit) unterrichten. Vorkenntnisse aus der Wahrscheinlichkeitsrechnung und der Begriff der Standardabweichung sind nicht zwingend erforderlich. Wir messen die Schwankungen durch Quantildifferenzen, deren Aussagekraft im Gegensatz zur Standardabweichung unmittelbar auf der Hand liegt.

2 Unterrichtsgang

1. S c h r i t t: Schätzen von Zufallsschwankungen

Um später untersuchen zu können, wie gut unsere intuitive Vorstellung von Zufallsschwankungen ist, beginnen wir mit einigen Schätzungen; diese können an geeigneter Stelle (Aufg. 6, Tab. 2) der Realität gegenübergestellt werden (Aufbau einer Erwartungshaltung, Sicherung eines Problemverständnisses). Ggf. kann man diesen Schritt auch überspringen und mit Schritt 2 beginnen.

Man bittet Schülerinnen und Schüler, sich vorzustellen, daß sie eine Münze 25mal werfen. Sie sollen aufschreiben, wie oft (auch prozentual gesehen) dabei »Kopf« gefallen sein könnte. Dieses Gedankenexperiment soll mehrfach wiederholt werden. SEGMEN schreibt: 12 (48%), 16 (64%), 19 (76%), 11 (44%), 8 (32%).

Die gleiche Frage beantwortet er für 100 Münzwürfe wie folgt: 48 (48%), 67 (67%), 27 (27%), 44 (44%), 62 (62%) und für 400 Münzwürfe: 227

(56,75%), 320 (80%), 130 (32,5%), 177 (44,25%), 210 (52,5%).

Die Schülerinnen und Schüler werden gebeten, für die ausgedachten relativen Häufigkeiten die kleinsten und die größten Werte anzugeben. Bei SEGMEN schwanken sie

für $n =$ 25 Versuche zwischen 32% und 76%,
für $n =$ 100 Versuche zwischen 27% und 67%,
für $n =$ 400 Versuche zwischen 32,5% und 80%.

SEGMENS Schätzungen drücken die Intuition aus, daß diese Schwankungen bei wachsender Versuchszahl etwa gleich groß bleiben werden. (Diese Intuition teilen erfahrungsgemäß nicht alle, aber viele Schülerinnen und Schüler, insbesondere solche, die ihre Zahlen ohne viel Nachdenken heruntergeschrieben haben.) Wir notieren einige der geschätzten Ergebnisse an der Tafel – und meist läßt eine engagierte Diskussion nicht lange auf sich warten. Die Frage, ob (bzw. wie) die Schwankungen relativer Häufigkeiten mit wachsender Versuchszahl abnehmen, ist für die Lerngruppe zum eigenen Problem geworden.

2. Schritt: Experimentelle Untersuchung von Zufallsschwankungen

Jeder wirft eine Münze 25mal und zählt aus, wie oft Kopf auftrat. Dieses Experiment wird einige Male wiederholt, so daß jeder auch eine relative Häufigkeit zu 100 Münzwürfen beisteuern kann. Wir fassen die Ergebnisse zusammen. Die letzte Spalte entstand hier durch Mittelwertbildung der ersten vier Spalten.

	(5 mal 25 Münzwürfe)					(100 Münzwürfe)
MICHELE:	0,52	0,36	0,48	0,24	0,68	0,40
MARCEL:	0,60	0,48	0,64	0,44	0,48	0,54
ANNA:	0,72	0,56	0,52	0,36	0,56	0,59
ANETTE:	0,52	0,56	0,40	0,48	0,64	0,49
RAINER:	0,52	0,52	0,36	0,40	0,44	0,45
KAI:	0,44	0,52	0,44	0,48	0,48	0,47
FRANK:	0,52	0,52	0,44	0,32	0,52	0,45
ANJA:	0,40	0,60	0,52	0,68	0,40	0,65
JENNET:	0,56	0,32	0,60	0,56	0,60	0,51
SEGMEN:	0,48	0,36	0,40	0,44	0,64	0,42

Schon der Vergleich der letzten Spalte mit den übrigen Daten läßt erahnen, daß die Zufallsschwankungen mit wachsender Versuchszahl (von 25 auf 100) abnehmen. SEGMENS Intuition trifft wohl nicht die Realität. Wir wollen die Abnahme quantitativ studieren.

3. Schritt: 96% und 68% – Schwankungsintervalle

Wir sortieren 50 relative Häufigkeiten (25 Münzwürfe) (siehe nächste Spalte oben).

Die relativen Häufigkeiten schwanken bei diesem Experiment zwischen 0,24 und 0,72 (Spannweite). Wir verabreden, die kleinste und die größte relative

1	MICHELE	Nr. 3	0,24		
2	JENNET	Nr. 2	0,32		
3	FRANK	Nr. 4	0,32		
4	MICHELE	Nr. 2	0,36		
5	ANNA	Nr. 4	0,36		
6	RAINER	Nr. 3	0,36		
7	SEGMEN	Nr. 2	0,36		
8	SEGMEN	Nr. 3	0,40		
9	ANJA	Nr. 1	0,40		
10	ANJA	Nr. 5	0,40		
11	RAINER	Nr. 4	0,40		
...				68%	96%
41	JENNET	Nr. 3	0,60		
42	JENNET	Nr. 5	0,60		
43	ANJA	Nr. 2	0,60		
44	MARCEL	Nr. 1	0,60		
45	ANETTE	Nr. 5	0,60		
46	MARCEL	Nr. 3	0,64		
47	SEGMEN	Nr. 5	0,64		
48	ANJA	Nr. 4	0,64		
49	MICHELE	Nr. 5	0,68		
50	ANNA	Nr. 1	0,72		

Häufigkeit als Ausreißer »wegzulassen«. Die restlichen 48 relativen Häufigkeiten (das sind 96% von 50) schwanken zwischen 0,32 und 0,68 (96%-Intervall). Lassen wir die acht kleinsten und die acht größten relativen Häufigkeiten weg, so liegen bei diesem Experiment die verbleibenden 34 (das sind 68% von 50) zwischen 0,40 und 0,60 (68%-Intervall). Die »Marken« 96% und 68% sind willkürlich gewählt. Andere Prozentsätze sind ebenso möglich, führen aber zu weniger einprägsamen Gesetzen. Den theoretischen Hintergrund der 1σ- und 2σ-Umgebungen und den zentralen Grenzwertsatz übergehen wir auf unserem Niveau experimenteller Untersuchungen.

4. Schritt: Computersimulation

Mit Computerhilfe[2] erhöhen wir nun die Anzahl der relativen Häufigkeiten (stets aus $n = 25$ Münzwürfen) von 50 auf 100, 200, ... (Vielfache von 50) und beobachten: Das 68%-Intervall bleibt (im Gegensatz zur Spannweite) davon unberührt, seine Breite schwankt nach wie vor um den Wert 0,2 (von 0,4 bis 0,6). Vervierfachen wir die Wurfzahl n von 25 auf 100, 400, ..., so halbiert sich die Breite schrittweise. Wir erhalten für die 68%-Intervalle Tabelle 1.

Wir merken uns folgende Faustregel:

Die Breite des 68%-Intervalls beträgt bei relativen Häufigkeiten aus n Münzwürfen $1/\sqrt{n}$. D. h. etwa 68% aller relati-

[2] Ein Simulationsprogramm (MS-DOS) zu den Schritten 4 und 8 kann gegen Einsendung einer formatierten Leerdiskette mit frankiertem Rückumschlag beim Verfasser angefordert werden.

n	von	bis	Breite
25	0,40	0,60	0,2 = 1/5
100	0,45	0,55	0,1 = 1/10
400	0,475	0,525	0,05 = 1/20
			$= 1/\sqrt{n}$

Tab. 1

ven Häufigkeiten liegen zwischen $0,5 - 1/(2\sqrt{n})$ und $0,5 + 1/(2\sqrt{n})$.

Das 96%-Intervall ist doppelt so groß. Insbesondere gilt: Man muß den Versuchsumfang vervierfachen, damit die relativen Häufigkeiten nur noch »halb so stark« schwanken.

5. Schritt: Übungen

Das $1/\sqrt{n}$-Gesetz muß durch kleinere Übungen gefestigt werden. Wir berechnen einerseits 68%- und 96%-Intervalle für relative und absolute Trefferhäufigkeiten, andererseits erforderliche Versuchsumfänge zu vorgegebenen Intervallbreiten. Dabei werden Algebra-Kenntnisse in inhaltlich sinnvollem Kontext aktiviert. Beispielaufgaben finden sich in Schritt 7.

Im Zusammenhang hiermit sollte man eine statistische Sprechweise für die inhaltlich erarbeiteten Zusammenhänge einführen:

Wir nennen ein experimentelles Ergebnis mit der Hypothese einer fairen Münze vereinbar, wenn die relative Häufigkeit innerhalb des 96%-Intervalls liegt. Andernfalls vermuten wir, daß die Münze wohl nicht fair ist.

Wie viele Deiner eingangs durchgeführten Schätzungen für relative Häufigkeiten aus 25, 100, 400 Münzwürfen könnten wir als Ergebnisse tatsächlicher Experimente noch akzeptieren? (Bei SEGMEN sind insgesamt 7 (46%) der 15 Schätzungen nicht akzeptabel. Hätte SEGMEN wirkliche Experimente – und keine Schätzungen – durchgeführt, sollten nur ca. 4% der Ergebnisse das Urteil »nicht akzeptabel« erhalten.)

6. Schritt: Statistischer Test

Es ist unverzichtbar, daß Schülerinnen und Schüler in eine tatsächliche Testsituation hineingestellt werden, deren Ausgang für sie von persönlicher Bedeutung ist. Konstruierte Beispiele sind »tödlich«. Hier ein Geschmackstest.

Testbogen: Koffein ist ein geschmackloses Aufputschmittel. Daher liegt die Hypothese nahe, daß man zwischen COKE mit (M) und ohne (O) Koffein nicht unterscheiden kann. Du wirst nach Beantwortung der Fragen a) bis c) 25 Kostproben erhalten. (25 numerierte, zufällig mit (M) bzw. (O) gefüllte Becher werden die Runde machen, jeder nimmt mit seinem Trinkhalm eine Probe und kreuzt in d) an.

a) In welchen Intervallen müßten relative und absolute Trefferhäufigkeit mit 96%iger Sicherheit liegen, wenn Du keinen Unterschied schmeckst, also Deine Antworten gleichsam zufällig (durch Werfen einer Münze) entstanden sind?

b) Beantworte die gleiche Frage für 250 Kostproben (die etwa durch Zusammenfassen der Ergebnisse von 10 Schülerinnen und Schülern entstehen.)

c) Welche Folgerung würdest Du ziehen, wenn Deine Trefferhäufigkeit
- »oberhalb« des 96%-Intervalls,
- »unterhalb« des 96%-Intervalls liegt?

(Im zweiten Fall ist man vermutlich ein Inversschmecker, COKE mit Koffein schmeckt der Testperson wie die koffeinfreie Variante und umgekehrt.)

d) Bewerte die 25 Geschmacksproben mit M bzw. O:

Probe Nr.:	1	2	3	4	5	1	2	3	4	5
Mein Tip:	O	O	O	M	M	O	M	M	O	O
Auswertung:	f	r	r	f	r	f	r	f	r	f
Probe Nr.:	1	2	3	4	5	1	2	3	4	5
Mein Tip:	O	O	M	M	M	O	M	O	M	O
Auswertung:	f	r	r	r	f	r	f	f	r	r
Probe Nr.:	1	2	3	4	5					
Mein Tip:	M	M	O	O	M					
Auswertung:	r	f	r	r	r					

Meine absolute Trefferhäufigkeit: 15.
Meine relative Trefferhäufigkeit: 60%.

Ist Dein Testergebnis mit der Hypothese vereinbar, daß Du keinen Unterschied schmeckst, oder liegt es im Schmeckerbereich bzw. im Inversschmeckerbereich? (In allen untersuchten Klassen waren mehrere »Schmecker«; nur sehr selten erhält man die Diagnose »Inversschmecker«.)

Der Variation dieser Testprobleme (gerade mit $p = 0,5$) sind keine Grenzen gesetzt. Kann man Magermilch von Vollmilch, kohlensäurefreies Mineralwasser von Leitungswasser, preiswerte No-Name Produkte von teuren Markenartikeln, Kunstleder von Leder, Naturfasern von Synthetik, CD-Spieler von Bandgeräten ... unterscheiden?

7. Schritt: Weitere Übungen und Klassenarbeit

Es folgen einige Aufgaben, die man auch für Klassenarbeiten verwenden kann.

Aufgabe 1: a) Was bedeutet die Aussage: »das 96%-Intervall der relativen Häufigkeit ist bei 400 Münzwürfen [0,45; 0,55]?«

b) Dir sagt jemand: »Ich habe in 400 Münzwürfen 241mal Kopf erhalten.« Wie beurteilst Du diese Aussage?

Aufgabe 2: a) Ulla hat bei 85 Geschmacksproben 53 Treffer erreicht. Ist das Ergebnis noch mit der Annahme vereinbar, daß Ulla zufällig angekreuzt hat, also keinen Unterschied schmeckt?

b) Bei welchen Trefferzahlen wird man die Annahme fallenlassen, daß Ulla bloß rät?