Das 'Eins durch Wurzel aus n'- Gesetz.
Einführung in statistisches Denken auf der Sekundarstufe I

von *Wolfgang Riemer*, Köln

Die folgende Unterrichtssequenz versucht (in Anlehnung an den neuen Lehrplan in NRW), an intuitive Vorerfahrungen anknüpfend, Schülerinnen und Schüler 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 ($x \to x^{-1/2}$) 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 elemantaren 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. 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.

Stochastik in der Schule 11 (1991), Nr. 3, 24-36.

1. Schritt: 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, Abb. 1) der Realität gegenübergestellt werden. (Aufbau einer Erwartungshaltung, Sicherung eines Problemverständnisses.) Gegebenenfalls kann man diesen Schritt auch überspringen und mit 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 *Gedanken*experiment 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 33% 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

Tab.1

Schon der Vergleich der letzten Spalte mit den übrigen Daten läßt erahnen, daß die Zufallsschwankungen mit wachsender Versuchszahl, hier 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 die 50 relativen Häufigkeiten aus den je 25 Münzwürfen, siehe Tabelle 2. Die relativen Häufigkeiten schwanken bei diesem Experiment zwischen 0.24 und 0.72. Wir verabreden, die kleinste und die größte relative 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.

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	

Tab.2

4. Schritt: Computersimulation

Mit Computerhilfe 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 folgende Tabelle:

n	von	bis	Breite
25	0.40	0.60	0.2 = 1/5
100	0.45	0.55	0.1 = 1/10
400	0.48	0.53	0.05 = 1/20
n			$1/\sqrt{n}$

Tab.3

Wer nicht selber programmieren oder in der Klasse mit Tabellenkalkulation arbeiten möchte, kann beim Autor ein Simulationsprogramm für *MS DOS* erhalten.

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 relativen Häufigkeiten liegen zwischen

$$0.5 - 1/(2\sqrt{n}) \quad \text{und} \quad 0.5 + 1/(2\sqrt{n}).$$

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. Wie berechnen einerseits 68%- und 96%-Intervalle für relative und absolute Trefferhäufigkeiten, andererseits erforderliche Verrsuchsumfänge 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 erabeiteten 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   1 2 3 4 5   1 2 3 4 5   1 2 3 4 5`

Mein Tip

`O O O M M   O M M O O   O O M M M   O M O M O   M M O O M`

Auswertung

`f r r f r   f r f r j   f f r r r   f r f f r   r r f 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 'Schmekker'; 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, 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?

Aufgabe 3: Wieviele Versuche muß man machen, damit die relative Trefferhäufigkeit mit 68%iger Sicherheit zwischen 0.46 und 0.54 liegt?

Aufgabe 4: Beurteile mit Hilfe des 'Eins durch Wurzel aus n'- Gesetzes, ob folgende Ergebnisse von echten Münzwürfen stammen können, oder ob sie eher ausgedacht sind:

a) 52 mal Kopf in 90 Würfen

b) 520 mal Kopf in 900 Würfen

c) Frank kritisiert die Aufgabenstellung: 'bei a) und b) ist das doch das gleiche'. Nimm Stellung zur Aussage von Frank!

Aufgabe 5: Für die Breite des 96%-Intervalls gilt $b=2/\sqrt{n}$. Begründe mit dieser Formel: Wenn man die Versuchszahl mit dem Faktor k vervielfacht, dann muß man die Breite durch $\sqrt{k}$ teilen, um das neue 96%-Intervall zu erhalten.

Aufgabe 6: Tab. 1 zeigt die von unserer Klasse (in der ersten Stunde) geschätzten relativen Häufigkeiten für 25/100/400 Münzwürfe in sortierter Reihenfolge.

a) Lies die Breite des Intervalls ab, in dem 96% unserer Schätzungen lagen für a_1) n=25 a_2) n=100 a_3) n=400. Vergleiche diese Breite mit dem Ergebnis des $1/\sqrt{n}$-Gesetzes.

b) Wieviel Prozent unserer Schätzungen liegen in dem durch das $1/\sqrt{n}$-Gesetz gelieferten 96%-Intervall für b_1) n=25 b_2) n=100 b_3) n=400.

c) Wie beurteilst Du unsere anfänglichen Schätzungen mit Deinem jetzigen Wissen?

Aufgabe 7: In Abb. 2 sind drei Funktionsgraphen zu sehen;

$$x \to 1/\sqrt{x}, \quad x \to 0.5+1/\sqrt{x}, \quad x \to 0.5-1/\sqrt{x}.$$

a) Überprüfe durch je zwei 'Stichproben', daß die Graphen richtig gezeichnet sind.

b) Welche Fragen lassen sich mit Hilfe der Graphen 1 und 2 durch Ablesen beantworten? Welche Bedeutung haben die x- und die y-Werte?

c) Jemand hat zwei Geschmackstests durch die Punkte A und B dargestellt; erläutere in eigenen Worten, wie die Versuche ausgegangen sind und welche Folgerungen man ziehen könnte.

	25 Versuche			100 Versuche			400 Versuche		
1	3	steffi	0,20	2	steffi	0,10	1	stephan	0,1800
2	1	martina	0,20	3	martina	0,10	1	bernd	0,1875
3	3	jola	0,28	1	bernd	0,17	4	steffi	0,2500
4	2	thao	0,28	4	gregor	0,17	2	robert	0,3075
5	2	gregor	0,28	2	simone	0,23	3	simone	0,3225
6	3	verena	0,32	4	stephan	0,23	3	bernd	0,3750
7	2	denise	0,32	1	steffi	0,25	1	steffi	0,3750
8	2	simone	0,36	1	thao	0,25	3	martina	0,3750
9	4	bernd	0,36	2	marc	0,28	1	gregor	0,3750
10	3	denise	0,36	1	stephan	0,30	2	thao	0,3875
11	4	stephan	0,36	2	bernd	0,31	4	silke	0,4300
12	2	silke	0,36	3	simone	0,32	3	denise	0,4475
13	2	marc	0,36	2	jola	0,32	1	verena	0,4500
14	1	steffi	0,40	1	gregor	0,32	3	silke	0,4500
15	2	stephan	0,40	4	verena	0,35	2	jola	0,4500
16	1	silke	0,40	2	denise	0,39	4	stephan	0,4700
17	3	gregor	0,40	1	silke	0,39	4	jola	0,4800
18	4	marc	0,44	1	verena	0,40	4	martina	0,4875
19	1	verena	0,48	2	robert	0,40	3	marc	0,4925
20	3	simone	0,48	3	denise	0,41	2	verena	0,4975
21	3	thao	0,48	4	martina	0,42	3	verena	0,5000
22	3	martina	0,48	3	verena	0,45	2	bernd	0,5000
23	2	bernd	0,52	3	silke	0,45	2	steffi	0,5000
24	4	denise	0,52	2	thao	0,45	3	stephan	0,5000
25	3	silke	0,52	1	simone	0,48	1	jola	0,5000
26	1	jola	0,52	4	bernd	0,48	4	marc	0,5075
27	4	thao	0,52	2	verena	0,49	1	thao	0,5125
28	3	marc	0,52	2	stephan	0,49	4	verena	0,5250
29	2	robert	0,52	3	steffi	0,50	4	denise	0,5275
30	1	denise	0,56	1	jola	0,50	2	silke	0,5325
31	4	silke	0,56	1	martina	0,50	1	denise	0,5500
32	1	bernd	0,60	4	silke	0,52	4	gregor	0,5500
33	2	steffi	0,60	2	gregor	0,53	3	thao	0,5550
34	1	stephan	0,60	3	marc	0,54	1	silke	0,5625
35	1	thao	0,60	1	denise	0,56	1	martina	0,5625
36	2	verena	0,64	4	jola	0,56	3	jola	0,5750
37	2	martina	0,64	3	thao	0,56	2	marc	0,5850
38	1	simone	0,68	4	denise	0,63	1	robert	0,5925
39	2	jola	0,68	2	silke	0,63	2	denise	0,5975
40	1	marc	0,68	1	marc	0,63	4	bernd	0,6500
41	1	gregor	0,68	4	simone	0,65	2	gregor	0,7000
42	4	simone	0,72	4	thao	0,65	2	simone	0,7450
43	4	verena	0,76	2	martina	0,65	4	thao	0,7525
44	3	bernd	0,76	3	bernd	0,68	1	marc	0,7525
45	4	gregor	0,76	3	jola	0,68	4	simone	0,7775
46	4	steffi	0,80	4	steffi	0,70	2	stephan	0,8000
47	3	stephan	0,80	3	stephan	0,70	2	martina	0,8000
48	4	martina	0,80	3	gregor	0,70	1	simone	0,8650
49	1	robert	0,80	1	robert	0,70	3	steffi	0,8750
50	4	jola	0,88	4	marc	0,81	3	gregor	0,8750

Tab. 4: Schätzungen relativer Häufigkeiten in der Klasse

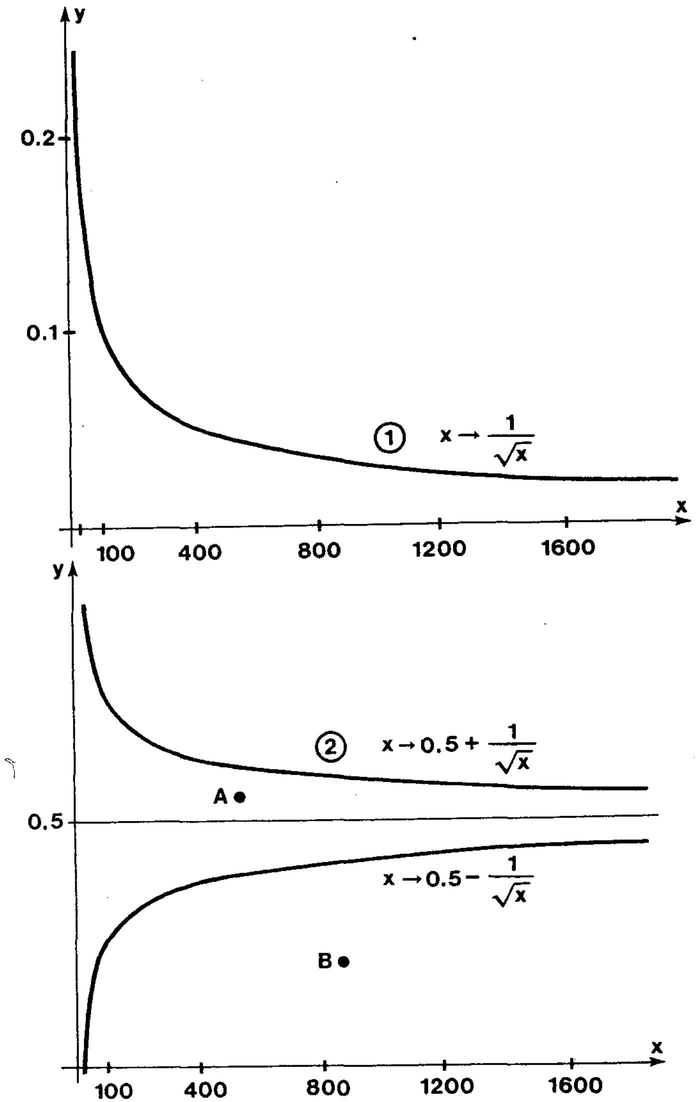

Fig. 1: Potenzfunktionen und Statistik

8. Schritt: Der Fall $p \neq 0.5$

Steht man nicht unter Zeitdruck, so läßt sich die Lernsequenz wie folgt fortsetzen: Die relativen Häufigkeiten schwanken nun von Versuch zu Versuch um die Wahrscheinlichkeit p, doch ist die Schwankungsbreite (68%- und 96%-Intervall) für $p \neq 0.5$ (etwas) kleiner als für $p=0.5$. Das kann man experimentell belegen und durch die Extremfälle $p=0$, $p=1$, in denen die Schwankungsbreite 0 werden muß, heuristisch absichern. Die 'alte' Formel $b=1/\sqrt{n}$ ist also für $p=0.5$ nur eine Abschätzung der Schwankungsbreite. Wenn man die exakte Formel $b=2\sqrt{p(1-p)}/\sqrt{n}$ erwähnen möchte, dann sollte man bestätigen, daß sie für $p=0.5$ und die genannten Extremfälle richtig arbeitet. Im Anfangsunterricht scheint es aber vertretbar, mit der Abschätzung zu arbeiten.

> Wir halten fest: Jede Wahrscheinlichkeit p hinterläßt im Experiment ihre 'Spuren' in Form (um p schwankender) relativer Häufigkeiten. Vervierfachen wir die Versuchszahl, so halbiert sich die Breite der Schwankungsintervalle - die Spuren werden also doppelt so deutlich erkennbar.

Aufgabe 8: Nach einer Umfrage (Stichprobengröße nicht angegeben) gab die GfK-Fernsehzuschauerforschung folgende Sehbeteiligungen an.

a) Für das WM Halbfinale Deutschland-England 49.9%. Ermittle das 68% Schwankungsintervall für die relative Sehbeteiligung, wenn Du folgende Stichprobengrößen unterstellst: a_1) $n=100$ a_2) $n=400$ a_3) $n=6500$.

b) Wie in a) für die Sendung 'Ein Fall für zwei' mit 17.8% Sehbeteiligung.

c) Bei der GfK ging man von 52 Mio Bundesbürgern aus, die Zugang zu einem Fernseher hattten. Wie viele Personen haben die Sendungen gesehen? Gib für die drei Stichprobengrößen auch die 68% Schwankungsintervalle der absoluten Häufigkeiten an.

d) Auf wieviele Stellen hinter dem Komma würdest Du die absolute Fernsehbeteiligung (in Mio Personen) angeben, wenn man die Stichprobengröße $n=6500$ benutzt? Die GfK verwendet diese Stichprobengröße und gibt 2 Stellen an, etwa 25.96 Mio.

9. Schritt: Tests, Konfidenzintervalle

Wir versetzen uns in die Lage einer Prüfungskommission, die die Aufgabe hat, Feinschmecker einzustellen. Den Kandidaten werden 250 Proben verabreicht. Wie soll das Einstellungskriterium aussehen? Es reicht sicher nicht, wenn wir den Kandidaten als Schmecker bezeichnen können, also nach Abschnitt 6 $p=0.5$ verwerfen. Denn schon bei einem sehr 'wackeligen' Geschmackssinn von $p=0.55$ wird das Testergebnis mit (mehr als) 96% iger Sicherheit im Bereich $0.55 \pm 1/\sqrt{250}$, also zwischen 0.518 und 0.582 liegen. Jemand, der nur ein bißchen schmeckt, würde nach diesem Kriterium gleich als Feinschmecker bezeichnet werden. Und das kann nicht Sinn der Prüfung sein. Wir einigen uns: Der Kandidat sollte schon die Trefferquote $p=0.8$ haben. Seine Ergebnisse müßten dann mit (mehr als) 96%iger Sicherheit zwischen 0.768 und 0.832 liegen. Natürlich stellen wir ihn auch ein, wenn seine Trefferquoten noch höher sind.

Übrigens: Die oben zitierte Gruppe von 10 zufällig ausgewählten Schülerinnen und Schülern hatte bei ihren insgesamt 250 Proben 142 Treffer, also die relative Häufigkeit 0.568. Sie liegt oberhalb des 96%-Intervalls von 0.5, (die Gruppe 'schmeckt'), aber weit unterhalb des 96%-Intervalls zu 0.8. Es handelt sich bei weitem nicht um eine Feinschmeckergruppe. Das Intervall der Trefferwahrscheinlichkeiten, die mit diesem Versuchsergebnis noch vereinbar sind, ist (kleiner als) [0.505;0.631]. Damit kann auf diesem elementaren Niveau eine Vorstellung von Konfidenzintervallen erarbeitet werden: Sie enthalten alle Wahrscheinlichkeiten, die mit einem Versuchsergebnis noch vereinbar sind.

Schlußbemerkung

a: Die Lernsequenz zeigt, wie man Grundgedanken beurteilender Statistik ohne Wahrscheinlichkeitsrechnung bearbeiten kann, wenn man das $1/\sqrt{n}$-Gesetz als Naturgesetz akzeptiert. Das Vorgehen ist zwar unüblich, hat aber den Vorteil, daß man ohne theoretischen Vorspann aus der Wahrscheinlichkeitsrechnung unmittelbar ins Zentrum statistischen Denkens und damit zu den für die Allgemeinbildung besonders wichtigen stochastischen Inhalten vordringt.

b: Es ist auch möglich, das $1/\sqrt{n}$-Gesetz in erster Linie als eine nicht triviale außermathematische Anwendung der Potenzfunktionen zu sehen, die sich harmonisch in eine sonst recht trockene Thematik einfügt.

c: In unserer Lernsequenz rückt die zentrale Bedeutung des Versuchsumfangs ins Zentrum des Bewußtseins und hilft so, gängige Fehlinterpretationen auszuräumen. Insbesondere wird klar, daß es nicht viel bedeutet, eine punktuelle Hypothese bei *sehr großem Versuchsumfang* verworfen zu haben. Die Situation gleicht dann einer Längenmessung mit Hilfe eines *Mikroskops*, bei der der Experimentator zum Schluß kommt, daß die Länge eines Stabes nicht genau 50 cm beträgt. (Aber wer denkt bei einer solchen Längenangabe schon an mikroskopische Genauigkeit?). Auch wird inhaltlich erfahrbar, daß das Verwerfen einer Hypothese ('schmeckt keinen Unterschied') nicht unbedingt als Bestätigung einer möglichen Alternative ('ist Feinschmecker') interpretiert werden darf.

Literatur

Riemer, W.: 1991, *Stochastische Probleme aus elementarer Sicht*, Mannhein: BI.

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