

Soundcheck: CD contra MP3

Ein Hörtest als Einstieg in die Stochastik

LERNGRUPPE: 7.–12. Schuljahr

IDEE: Ab welcher Trefferzahl im Multiple-Choice-Test gilt man als Profi-Hörer? Um ein Entscheidungskriterium zu finden, wird simuliert und Theorie entwickelt.

ZEITBEDARF: 4–8 Unterrichtsstunden

WEITERES MATERIAL: Musikdateien und Excel-Arbeitsblätter sind unter www.mathematik-lehren.de als Download zu ML153 erhältlich.

Stochastik setzt sich aus der Wahrscheinlichkeitsrechnung, beschreibender und beurteilender Statistik zusammen. Spannend wird es, wenn es um authentische Probleme mit realen Daten geht. Der hier vorgestellte Hörtest bietet ein ideales Experimentierfeld, das die drei Teilgebiete der Stochastik zusammenwachsen lässt. Ähnlich spannend sind auch Geschmackstests (Cola, Chips, Schokolade; s. Riemer 1997) –

aber Musik krümelt und klebt nicht, die ganze Klasse kann simultan testen und MP3-Musik genießt im „IPOD-Zeitalter“ einen hohen „Sex-Appeal“.

Vor der Auswertung des Tests wird mit Simulationen (ohne Wahrscheinlichkeitsrechnung!) die Brücke von der beschreibenden zur beurteilenden Statistik geschlagen und in einer vertiefenden Schleife die Wahrscheinlichkeitsrechnung (Pfadregel, Binomialverteilung) eingebaut.

Stochastik unterrichten: Paradigmen

Stochastik lebt von Experimenten, die echte Fragen beantworten sollen. Vor Auswertung der Experimente sollten Schülerinnen und Schüler aber spekulieren, Hypothesen aufstellen, mit diesen Hypothesen eigene Prognosen machen und diese an der Realität messen.

Simulationen, die Pfadregel und die Binomialverteilung werden dabei als Werkzeuge genutzt bzw. erarbeitet.

Wenn die Versuchsauswertung erst nach der Theoriephase erfolgt, in der man (z. B. mit Simulationen) erforscht, wann man „angenommene – hypothetische – Wahrscheinlichkeiten“ anzweifeln (verwerfen) sollte, ist man mitten in der beurteilenden Statistik. Wenn man die Versuche direkt nach der Durchführung auswertet, landet man in der beschreibenden Statistik. Der Unterricht wird, gemessen an den potentiellen Möglichkeiten, „öde“. Er verliert Tiefgang, es schwindet der Zwang zum Nachdenken, zum Hypothesen bilden und zum Modellieren.

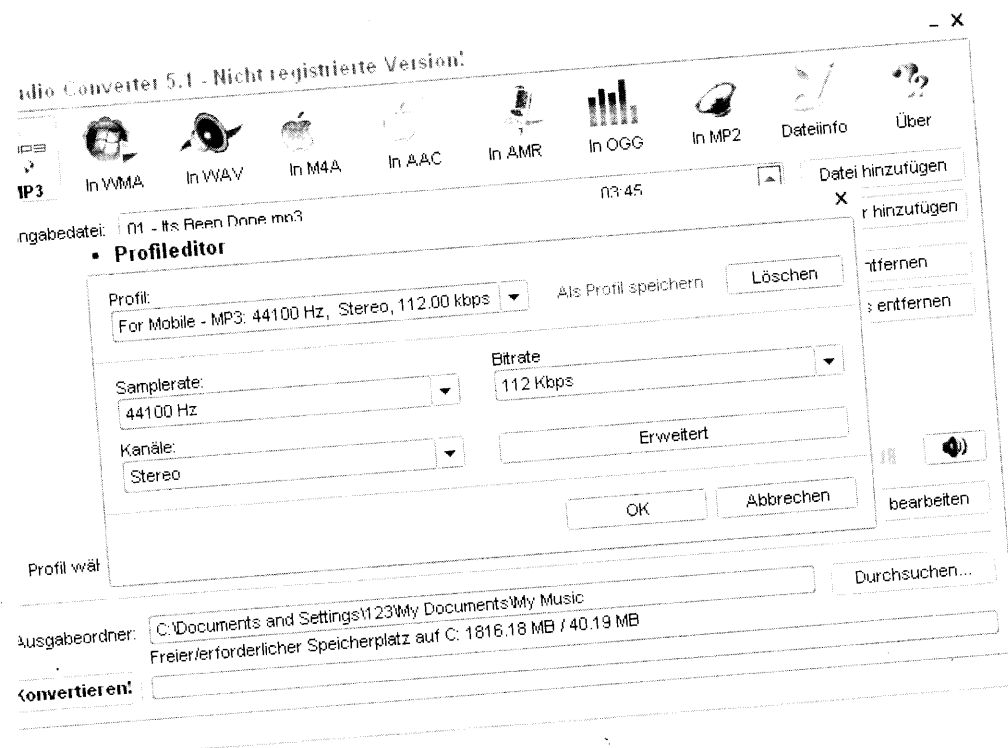
Wie funktioniert der Hörtest?

Beim Hörtest wird ein kurzer Ausschnitt eines Musikstücks 20 mal vorgespielt – in vier verschiedenen Qualitätsstufen. Die Testpersonen kreuzen an, um welche Qualität es sich ihrer Meinung nach handelt.

Vorbereitung

Man kann mit den vorbereiteten Musikstücken arbeiten oder ein Lied von einer CD auf den Computer „grabben“ und sowohl als wav-Datei in CD-Qualität wie auch als komprimierte mp3-Datei abspeichern. Als Kompressions-einstellung bieten sich konstante Bitraten an: 64 bit/s (spart Speicherplatz), 128 bit/s (Standardqualität) und 192 bit/s (hochwertige Qualität).

Musik aufs Handy – kein Problem. Nur was stelle ich ein?



Der Versuchsleiter „würfelt“ für ein Musikstück eine Playliste mit 20 Qualitätsstufen aus: Bei „1“ wird die wav-Datei, bei „2“ die 192er-mp3-Datei, bei „3“ die 128er-mp3-Datei, bei „4“ die 64er-mp3-Datei in die Liste eingereiht. Bei „5“ und „6“ wird weiter gewürfelt. Für die Zuhörer ist nicht bekannt, welche Qualitäten in welcher Reihenfolge in der Playliste stehen. Es kann sogar passieren, dass nicht alle Qualitätsstufen vorkommen, denn jede Stufe wird beim Auswürfeln nur mit der Wahrscheinlichkeit $\frac{1}{4}$ ausgewählt. Das macht das Hörexperiment anspruchsvoll.

Durchführung

Der Test wird in einem Musikraum mit guten Boxen oder in einem HiFi-Studio (z.B. in einem entsprechenden Kaufhaus) durchgeführt. Die Testbögen (s. Abb. 1) werden ausgefüllt und in einem Umschlag versiegelt. (Mit der Datei *ml_153_Hörtest_Testbogen.xls* können Sie leicht Kopiervorlagen erstellen.) Vor der Auswertung der Bögen wird erst einmal simuliert und nachgedacht („Spannungsbogen“ nutzen).

Simulation des Hörtests

Wie viele Treffer erhält jemand, der keine Unterschiede hört, der also nur zufällig ankreuzt („Nullhörer“, „Zufällig-ankreuzer“) und bei dem alle Treffer Zufallstreffer sind? Ab welcher Trefferzahl wird man einem Testkandidaten die Fähigkeit zusprechen, dass er Unterschiede „hört“? (Statistiker bezweifeln oder verwerfen dann die Nullhörer-Hypothese.) Die Antwort findet man durch Simulieren. Die Datei *ml_153_Hörtest_Simulation.xls* kann als Simulationsvorlage dienen (vgl. Abb. 2).

Wieder gibt es einen „Versuchsleiter“. Die anderen sind Kandidaten und würfeln als „Nullhörer“ die „gehörten“ Qualitätsstufen zwischen 1 und 4 mit einem richtigen Würfel aus und tragen ihr Ergebnis („gehörte Qualitätsstufe“) in die Spalte „Kandidat“ ein. Unabhängig davon würfelt der Leiter die „Playliste“ und trägt sie bei „Leiter“ ein. Die Playliste wird vorgelesen und jeder notiert die Anzahl seiner Treffer. Die Simulation wird fünfmal wiederholt, und die Ergebnisse werden in Balken-

diagrammen oder Boxplots visualisiert (s. Abb. 3). Zur Ergänzung bieten sich Grafikrechner oder eine Computersimulation an (*ml_153_Hörtest_Computer-simulation.xls*).

Auch eine effektive Auswertung im Plenum ist möglich. Dann trägt jeder Kandidat seine Trefferzahlen und Abstandssummen in Strichlisten an der Tafel ein. Als Ergebnis der Simulation weiß man dann ungefähr, in welchen Bereichen Trefferzahlen und Abstandssummen normalerweise liegen, wenn man zufällig ankreuzt.

Entscheidungskriterium finden:

Wer ist ein MP3-Qualitätshörer?

Das Balkendiagramm und der Boxplot aus Abb. 3 zeigen, dass die Trefferzahl von Nullhörern bei der Hälfte der Versuchsteilnehmer zwischen 4 und 7 liegt. Wie hoch müsste – nach einem Blick auf weitere Simulationen mit einem grafikfähigen Taschenrechner oder einer Tabellenkalkulation – bei 20 Hörproben die Trefferzahl sein, damit man „guten Gewissens“ sagen kann, dass ein Kandidat „wirklich besser ist“ als ein Nullhörer?

Schülerinnen und Schüler einigen sich – genau wie Wissenschaftler – nach intensiver Diskussion auf ein Entscheidungskriterium nach folgendem Muster: Fast immer (mit ca. 93%-iger Wahrscheinlichkeit) liegen bei Nullhören die Trefferzahlen X im Bereich $2 \leq X \leq 8$. Wenn die Trefferzahl über 8 liegt, bezweifeln wir, dass es sich um einen Nullhörer handelt. Da aber auch Trefferzahlen in der Nähe von Null bei Nullhörern „praktisch nie“ vorkommen, glauben wir, dass es sich in diesen Fällen um „Invershörer“ handelt. Das sind Leute, die zwar deutliche Unterschiede zwischen den Musikqualitäts-

	Musikstück1				Auswertung		
	1 = CD	2 = 192	3 = 128	4 = 64	ist richtig/falsch (-0/1)	Abstand	
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
Treffer							
Abstandssumme							

Abb. 1: Testbogen für eine Playliste (später wird rechts der Abstand zwischen gehörter und realer Qualitätsstufe eingetragen)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Simulationsblatt zum Würfeln																
2	Muster																
	Simulation A			Simulation B			Simu										
	Kand	Leite	Abst	Treff	Kand	Leite	Abst	Treff	Kand	Leite	Abst	Treff	Kand				
3																	
4	1	4	3	1	0	1			1				1				
5	2	1	4	3	0	2			2				2				
6	3	3	2	1	0	3			3				3				
7	4	1	1	0	1	4			4				4				
8	5	4	4	0	1	5			5				5				
9	6	2	2	0	1	6			6				6				
10	7	3	3	0	1	7			7				7				
11	8	1	1	0	1	8			8				8				
12	9	3	1	2	0	9			9				9				
13	10	1	4	3	0	10			10				10				
14	11	1	3	2	0	11			11				11				
15	12	2	4	2	0	12			12				12				
16	13	1	1	0	1	13			13				13				
17	14	1	2	1	0	14			14				14				
18	15	4	2	2	0	15			15				15				
19	16	2	3	1	0	16			16				16				
20	17	3	4	1	0	17			17				17				
21	18	2	2	0	1	18			18				18				
22	19	2	3	1	0	19			19				19				
23	20	3	3	0	1	20			20				20				
24						20	8										

Abb. 2: Das Arbeitsblatt hat Platz für fünf Simulationen

ten hören, aber sie systematisch falsch zuordnen. MP3-Musik klingt in ihren Ohren wie CD-Musik. Unsere Entscheidungsregel: Bei 9 oder mehr Treffern giltst du als „MP3-Qualitätshörer“. Bei 0 oder 1 Treffern giltst du als „MP3-Invershörer“.

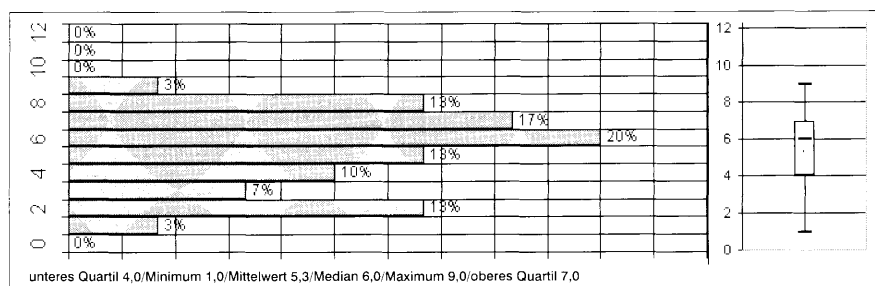


Abb. 3: Auswertung einer Excel-Simulation, die Hälfte der 30 Tester hatte zwischen 4 und 7 Treffer

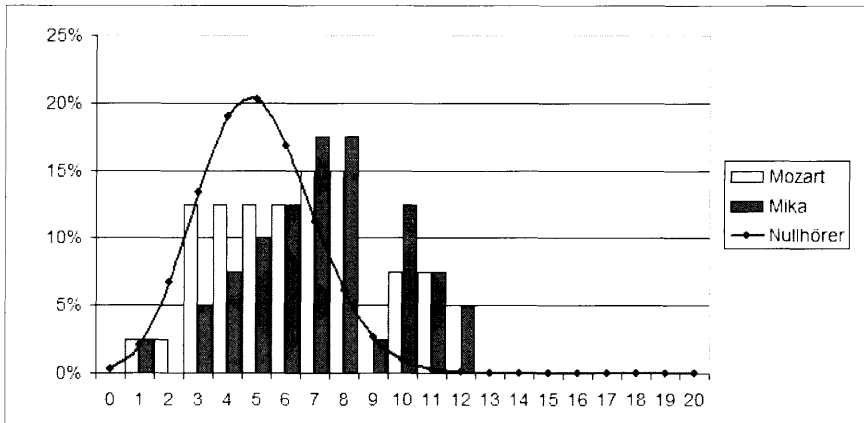


Abb. 4: Relative Häufigkeiten und Wahrscheinlichkeiten beim Hörtest mit zwei Musikstücken

		angekreuzt			
		1	2	3	4
tatsächlich	1	0	1	2	3
	2	1	0	1	2
	3	2	1	0	1
	4	3	2	1	0
Abst.	Wk.	0,25	0,375	0,25	0,125

Abb. 5: Verteilung der Abstände, wenn alle Qualitätsstufen mit gleicher Wahrscheinlichkeit angekreuzt werden

Auswertung des Hörtests

Die versiegelten Testbögen werden ausgeteilt. Der Versuchsleiter gibt die wahren Qualitätsstufen bekannt. Wer mit seiner Trefferzahl deutlich über der Trefferzahl eines Nullhörers liegt, kann für sich die Hypothese verwerfen, dass er selber ein „Nullhörer“ ist.

Alle Testpersonen im Blick: Binomialverteilungen

Die Frage nach den Wahrscheinlichkeiten, mit denen ein zufällig ankreuzender Nullhörer 0, 1, ..., 20 Treffer erhält, ist das Tor zum Einstieg in die Wahrscheinlichkeitsrechnung. Die Antwort gibt die Binomialverteilung: P(k Treffer in 20 Versuchen)

$$= \binom{20}{k} \cdot \left(\frac{1}{4}\right)^k \cdot \left(\frac{3}{4}\right)^{20-k}$$

So erhält man zum Beispiel 5 Treffer mit der Wahrscheinlichkeit

$$\binom{20}{5} \cdot \left(\frac{1}{4}\right)^5 \cdot \left(\frac{3}{4}\right)^{15} = 20,2\%$$

Einen Vergleich dieser Wahrscheinlichkeiten mit den in einer Lerngrup-

pe erzielten relativen Trefferhäufigkeiten zeigt **Abb. 4** (im Test eine Arie aus Mozarts Figaro und „Grace Kelley“ des Sängers Mika). Die Häufigkeitsverteilungen sind gegenüber den „binomialen Nullhörer-Wahrscheinlichkeiten“ tatsächlich zugunsten höherer Trefferzahlen verschoben: Bei „Mozart“ konnten 4, bei „Mika“ 10 von jeweils 40 Testern die Nullhörer-Hypothese verwerfen.

Durch Zusammenfassen von Qualitätsstufen kann man im Rahmen von Übungsaufgaben die Versuchsergebnisse auch Binomialmodellen mit anderen Trefferwahrscheinlichkeiten ($p = \frac{1}{2}$ oder $p = \frac{1}{3}$) gegenüberstellen. Wer im Anschluss an die Pfadregel einen glatten Einstieg in die Binomialverteilung etwa schon in einer Klasse 7/8 sucht, sei auf (Riemer 2007) verwiesen.

Wie sehr liegt man „daneben“? Abstandssummen helfen weiter

Pfiffige Schülerinnen und Schüler sind mitunter mit dem Zählen der Treffer als Kriterium für gutes Gehör nicht zufrieden. Sie schlagen vor, auch die Abstände zwischen angekreuzter und tatsächlicher Qualitätsstufe zum Kriterium zu machen. So liegt ein Kandidat, der wav-Musik (1) für 192er-Qualität (2) hält, näher an der Wirklichkeit als jemand, der sie für 64er-Qualität (4) hält. Der Qualitätsstufenabstand ist im ersten Fall 1, im zweiten Fall 3. Bei 20 Musikstücken kann die Abstandssumme zwischen D = 0 (alle Musikstücke richtig erkannt) und D = 60 (immer 3 Qualitätsstufen daneben) schwanken.

Um dieses Testergebnis bewerten zu können, müsste man wissen, wie die Abstandssummen bei Nullhörern verteilt sind. Dazu werden die Abstandssummen simulierter Nullhörer in einem Balkendiagramm visualisiert und mit der Realität (Abstandssummen aus der Testauswertung) verglichen. Für eine fundierte Beurteilung braucht man aber die Wahrscheinlichkeitsverteilung der Abstandssumme.

Wahrscheinlichkeitsverteilung der Abstandssumme

Um die Wahrscheinlichkeitsverteilung zu berechnen, ist man auf eine rekursive Berechnung mit Tabellenkalkulation angewiesen, da es keine zur Binomialverteilung analoge Wahrscheinlichkeitsformel gibt.

Verteilung des Qualitätsabstandes bei einem Musikstück

Wenn der Kandidat alle vier Qualitätsstufen mit gleicher Wahrscheinlichkeit zufällig ankreuzt, gibt es 16 gleichwahrscheinliche Kombinationen, von denen z. B. vier zum Abstand 0 führen, sechs zum Abstand 1; vier zum Abstand 3 und zwei zum Abstand 4. Damit ergibt sich die Verteilung aus **Abb. 5**.

Schrittweise Berechnung der Abstandssumme nach 2, 3 ... Liedern

Die Verteilung des Qualitätsabstandes wird in **Abb. 6** dargestellt. In Zeile 3 stehen die berechneten Wahrscheinlichkeiten nach einem („dem ersten“) Musikstück (*ml_153_Hörtest_Rekursion.xls*). Die Wahrscheinlichkeit, dass man nach dem zweiten Musikstück insgesamt die Abstandssumme 4 hat, berechnet man so:

- entweder hatte man nach einem Schritt den Abstand 3 und im zweiten Schritt kommt der Abstand 1 hinzu (12,5%*37,5%, H3*F3),
- oder man hatte den Abstand 2 und es kommt der Abstand 2 hinzu (25,0%*25,0%, G3*G3),
- oder man hatte den Abstand 1 und es kommt der Abstand 3 hinzu (37,5%*12,5%; F3*H3).

Diese Werte werden addiert und mit der in der Kopfzeile sichtbaren Formel

in der markierten Zelle I4 „hinterlegt“ Man erhält also nach zwei Schritten die Abstandssumme 4 mit der Wahrscheinlichkeit 15,6%. (Die Wahrscheinlichkeitsverteilung der Abstandssumme nach 20 Liedern zeigt sich in **Abb. 7**.)

Mit einer Wahrscheinlichkeit von ca. 95% liegt die Abstandssumme im Bereich $[25 - 8; 25 + 8] = [17; 33]$. Wenn wir Kandidaten bei einer Abstandssummen unter 17 als „Hörer“ und über 33 als „Invershörer“ bezeichnen, dann haben wir in der Testgruppe eine Reihe von „Hörern“ und keinen „Invershörer“. Die Häufigkeitsverteilung der Abstandssumme ist verglichen mit der Wahrscheinlichkeitsverteilung von Nullhörern deutlich in Richtung kleinerer Abstände verschoben. Die Testgruppe schneidet insgesamt besser ab als eine gleich große Nullhörerguppe.

„Systemankreuzer“ und die Tendenz zur Mitte

Ob man die Testgröße „Abstandssumme“ durch *strategisches Ankreuzen* verbessern (d.h. verkleinern) kann? Wie sähe ihre Wahrscheinlichkeitsverteilung aus, wenn man als Nullhörer die Kreuzchen nicht zufällig setzt, sondern systematisch immer bei wav (1) oder immer bei MP3-192 (2)? Dazu betrachtet man im oberen Teil von **Abb. 5** nur die erste bzw. nur die zweite Spalte. Für ein Musikstück verändert sich die Wahrscheinlichkeitsverteilung des Abstandes: Bei einem Wav(1)-Ankreuzer beträgt die Wahrscheinlichkeit für jeden Abstand 0,25. Beim 128 (2)-Ankreuzer ist $WK(0) = 0,25$, $WK(1) = 0,5$; $WK(2) = 0,25$ und $WK(3) = 0$.

Um die Wahrscheinlichkeitsverteilung der Abstandssumme mit der Tabellenkalkulation zu berechnen, tauscht man die vier Wahrscheinlichkeiten für das erste Musikstück entsprechend aus. **Abb. 7** zeigt das Ergebnis – auch im Vergleich zur realen relativen Häufigkeitsverteilung der Testgruppe. Die 192-Systemankreuzer werden am besten abschneiden, weil sie deutlich geringere Abstandssummen erreichen.

Man erkennt (auch an den Kenngrößen Erwartungswert und Standardabweichung): Durch systematisches An-

	A	E	F	G	H	I	J	K	L	M	N	O	F
1	rekursive Berechnung der Wahrscheinlichkeiten für die Abweichungssummen bei Nullhörern												
2		0	1	2	3	4	5	6	7	8	9	10	
3	1	25,0%	37,5%	25,0%	12,5%								
4	2	6,3%	18,8%	26,6%	25,0%	15,6%	6,3%	1,6%	0,0%	0,0%	0,0%	0,0%	0,0%
5	3	1,6%	7,0%	15,2%	21,7%	22,3%	17,0%	9,8%	4,1%	1,2%	0,2%	0,0%	0,0%
6	4	0,4%	2,3%	6,8%	13,1%	18,4%	19,9%	17,1%	11,7%	6,4%	2,7%	0,9%	0,0%
7	5	0,1%	0,7%	2,7%	6,5%	11,5%	16,0%	18,0%	16,6%	12,8%	8,1%	4,3%	1,0%
8	6	0,0%	0,2%	1,0%	2,8%	6,1%	10,3%	14,2%	16,3%	15,9%	13,2%	9,4%	5,0%
9	7	0,0%	0,1%	0,3%	1,1%	2,8%	5,7%	9,3%	12,7%	14,9%	15,1%	13,3%	10,0%

Abb. 6: Rekursive Berechnung der Wahrscheinlichkeiten für die Abstandssumme bei Nullhörern

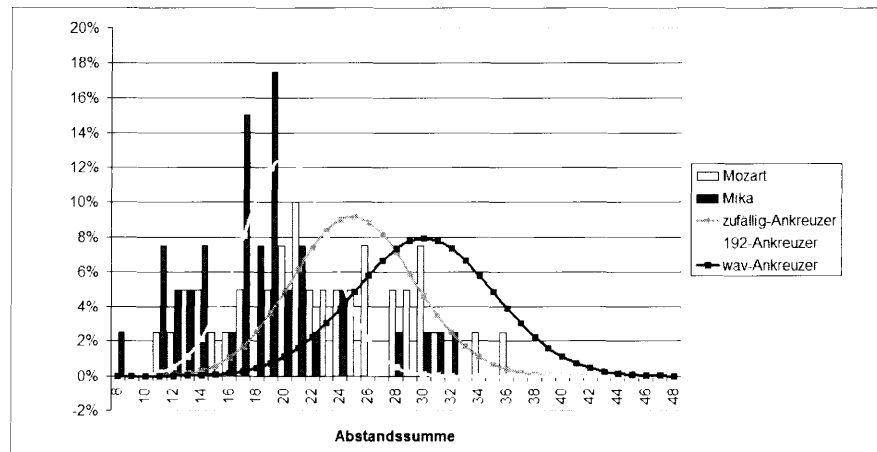


Abb. 7: Die realen relativen Häufigkeitsverteilungen passen am besten zu „192-Ankreuzern“

kreuzen der mittleren Qualitäten (192 oder 128) kann auch ein Nullhörer seine Abstandssumme verkleinern.

Auch bei den realen Testpersonen ist in der Tat eine unbewusste „Tendenz zu Mitte“ erkennbar. Auf die Trefferzahl wirkt sich diese Strategie, „eher die Mitte anzukreuzen“, nicht aus. Deswegen ist die Trefferzahl entgegen der Intuition wohl doch die bessere Testgröße.

Zentraler Grenzwertsatz

Aus dem Blickwinkel mühsamer Berechnungen zu Summenverteilungen, die im „Vorcomputerzeitalter“ gar nicht möglich war, umgibt den universell gültigen zentralen Grenzwertsatz ein „wohlverdienter Glorienschein“. Egal, wie ein Summand (hier der Qualitätsabstand zwischen gehörter und angekreuzter Qualität) verteilt ist, es kommt nur auf den Erwartungswert μ_1 und die Standardabweichung σ_1 an: Die Summe ist bei hinreichend großer Summandenzahl (hier 20) stets näherungs-

weise normalverteilt mit dem Erwartungswert $\mu_{20} = 20 \mu_1$ und $\sigma_{20} = \sqrt{20} \sigma_1$.

Es gilt die Sigmaregel: Im 1σ -Intervall um den Erwartungswert μ liegen ca. 68% aller Daten und etwa 95% aller Daten liegen im $1,96\sigma$ -Intervall. Für den zufällig ankreuzenden Nullhörer mit $\mu_{20} = 20 \cdot 1,25 = 25$ und $\sigma_{20} = \sqrt{20} \cdot 0,9375 = 4,193$ müsste also die Abstandssumme mit ca. 95% Wahrscheinlichkeit im entsprechenden Intervall $[25 - 1,96 \cdot 4,193; 25 + 1,96 \cdot 4,193] = [16,78; 33,22]$ liegen. Das passt hervorragend zu dem oben beschriebenen Intervall $[17; 33]$. Der Hörtest bietet auch hier authentisches Datenmaterial.

Literatur

- Riemer, W.: Geschmackstests: Spannende und verbindende Experimente – In: mathematik lehren Heft 85, Friedrich Verlag, 1997.
- Riemer, W.: Lambacher Schweizer Klasse 7. – Klett 734471, Stuttgart 2007, S. 62 f.
- Riemer, W.: Lambacher Schweizer Klasse 8. – Klett 734481, Stuttgart 2008.

Hinweis

Es gibt kostenlose Programme zum „graben“: z. B. eac oder audiograbber. Weitere Hörbeispiele und die Excel-Dateien können abgerufen werden unter: www.riemer-koeln.de