Bestell-Nr. 524214

MU

DER MATHEMATIK-UNTERRICHT

Blabla blabla blabla blabla p=0,4 blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla 800 blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla unter 40 % blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla höchstens 4 % blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla blabla

FRIEDRICH

Schickt die statistische Signifikanz in den Ruhestand!

Jahrgang 66 · Heft 4 · Juli 2020

Jahrgang 66 · Heft 4–2020

DER MATHEMATIK-UNTERRICHT

Beiträge zu seiner fachlichen und fachdidaktischen Gestaltung

Schwerpunkt **Schickt die statistische Signifikanz in den Ruhestand!**
Wolfgang Riemer

Adressen der Autoren

Norbert Henze
Henze@kit.edu

Thomas Hotz
thomas.hotz@tu-ilmenau.de

Wolfgang Riemer
w.riemer@arcor.de

Birgit Skorsetz
Birgit.Skorsetz@thillm.de

Reimund Vehling
vehling@icloud.com

BEITRÄGE

WOLFGANG RIEMER, REIMUND VEHLING

Prognose- und Konfidenzintervalle: beurteilende Statistik mit Sinn und Verstand

1 Den Zufall erforschen – statt ihn wegzuwünschen!

Der frequentistische Wahrscheinlichkeitsbegriff hat neben dem LAPLACEschen im Schulunterricht eine lange Tradition. FREUDENTHALs Notiz aus dem Jahr 1972 (vgl. **Abb. 1**) ist aktuell wie eh und je. Man deutet Wahrscheinlichkeit als eine objektive Größe, die man als Grenzwert einer unendlichen Folge relativer Häufigkeiten erhält[1]. Dahinter steckt das starke Gesetz der großen Zahlen[2], das besagt: Bei einer BERNOULLIkette konvergiert die Folge der relativen Häufigkeiten mit Wahrscheinlichkeit 1 gegen die Trefferwahrscheinlichkeit p.

Frequentistische Wahrscheinlichkeit
The story about tossing a coin with the happy result of a fair distribution of heads and tails in the long run has been the custom for quite a long time.

What is new about it, is that the story is dramatized and acted out – I mean, by the author of the textbook. Maybe even the teachers or the students are expected to try out this experiment – following the highly encouraging examples given by the textbook authors.

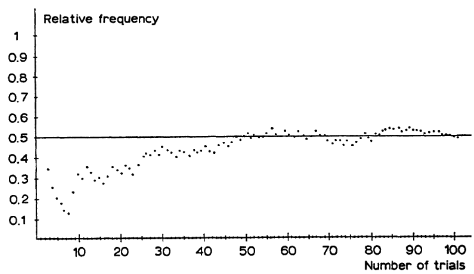


Abb. 1 Wenn man nach R. v. MISES Wahrscheinlichkeit als Grenzwert relativer Häufigkeiten deutet, dann sind Zufallsschwankungen unerwünscht.

The experiments reported in those textbooks attempt to prove the 'empirical law of large numbers'. Notwithstanding their great diversity all of them show the essential features displayed by the relative frequencies, that is *oscillation around, and a quick convergence towards,* $1/2$. In general, after 100 trials the value $1/2$ is practically reached, though some of the more optimistic textbook authors promise this result after no more than 50 trials. In some trials series the quotient 1/2 is suddenly reached at $n=100$, but in the majority of examples the quotients have been approaching close to 1/2 for a while before the 100th trial. FREUDENTHAL [1972]

1 Bevor KOLMOGOROFF Wahrscheinlichkeit axiomatisch fasste, verwickelt sich R. v. MISES beim Versuch, sie über Grenzwerte zu definieren, in Widersprüche. Vgl. HUMENBERGER [2019] zur Wortschöpfung „empirische Wahrscheinlichkeit".

2 Das ist ein tiefliegender mathematischer Satz, für dessen Beweis man ein Wahrscheinlichkeitsmaß auf dem Grundraum Ω aller Zahlenfolgen konstruieren muss. Die Wahrscheinlichkeit, dass eine unendliche Folge konvergiert, lässt sich nicht mehr frequentistisch deuten.

Leider hat die frequentistische Sicht den Haken, dass es keine unendlich langen Versuchsreihen gibt, mit denen man eine unbekannte Wahrscheinlichkeit genau bestimmen könnte.

Und wenn man es trotzdem versucht, stören die Zufallsschwankungen enorm. Man wünscht sie weg und vertreibt – wie **Abb. 1** belegt – den Zufall aus dem Stochastikunterricht. Zielführender ist es, schon in der Sekundarstufe I einen *hypothetisch-prognostischen* Wahrscheinlichkeitsbegriff (vgl. [RIEMER 2019]) zu pflegen und Wahrscheinlichkeiten als vom Menschen gesetzte *Modelle* zu begreifen, die

- relative Häufigkeiten prognostizieren,
- die Realität nur mehr oder weniger gut und nie ganz genau beschreiben,
- im Modellbildungskreislauf beständig nachgebessert werden müssen[3],
- nicht die Konvergenz relativer Häufigkeiten *in einer immer länger werdenden* Versuchsserie fokussieren, sondern die mit immer größer werdendem Versuchsumfang abnehmenden Zufallsschwankungen in *vielen endlich langen* Versuchsserien[4].

Damit ist man bei Prognoseintervallen, dem $1/\sqrt{n}$ -Gesetz und dem „*Konzept des Bezweifelns*" von Hypothesen, einer präformalen Version zweiseitiger Hypothesentests. Sie bilden die Basis für die Konstruktion von Konfidenzintervallen, die (anders als Testgrößen) über ihre Länge den Einfluss des Stichprobenumfangs verraten und *über die Idee der Messungenauigkeit eine Brücke zwischen Geometrie und Stochastik schlagen*.

Es gibt ausgezeichnete systematisch aufgebaute Lehrgänge zum Thema, z. B. den fachlichen Beitrag HENZE [2020] in diesem Heft und solide Schulbücher wie z. B. BRANDT u. a. [2018] oder LERGENMÜLLER u. a. [2012]. Deswegen fassen wir uns bei der Sachanalyse in 2.1 und 3.1 kurz und stellen praxiserprobte und didaktisch durchaus „pfiffige" Zugänge in den Vordergrund, wie sie sich in Schulbüchern nicht abbilden lassen.

2 Prognoseintervalle, das $1/\sqrt{n}$-Gesetz und das Konzept des Bezweifelns

2.1 Worum geht es?

Ein 95%-Prognoseintervall zu einer Trefferwahrscheinlichkeit p ist ein Intervall, in das die zufällige Trefferanzahl X bzw. die zufällige relative Trefferhäufigkeit $H = X/n$ bei einem *bevorstehenden* BERNOULLIexperiment gegebener Länge n mit einer Mindestwahrscheinlichkeit $95\,\%$ fällt[5]. Wie der Name „Prognose" signalisiert, schauen Prognoseintervalle – genau wie Wahrscheinlichkeiten – in die Zukunft. Man schließt mit ihnen vom Modell (der Wahrscheinlichkeit) auf die Wirklichkeit (die relative Häufigkeit).

Die Arbeit mit Prognoseintervallen

- vertieft den prognostischen Aspekt des Wahrscheinlichkeitsbegriffs,
- bietet eine griffige und nützliche Interpretation des „empirischen Gesetzes der großen Zahlen", die sich am schwachen, nicht am starken Gesetz der großen Zahlen orientiert und
- deckt über das *Konzept des Bezweifelns* die Anforderungen der Bildungsstandards hinsichtlich beurteilender Statik auf Grundkursniveau ab.

---

3 L. SACHS [1999] formuliert prägnant: „Aus Erfahrung wird Erwartung."

4 In der Schule ist es viel sinnvoller, das oft zitierte *empirische* Gesetz der großen Zahlen im Sinne des *schwachen*, nicht des *starken* Gesetzes der großen Zahlen zu interpretieren.

5 Absolute und relative Häufigkeiten sind Zufallsgrößen, die man mit Großbuchstaben (X, H) schreibt.

Das $1/\sqrt{n}$ -Gesetz besagt: Bei gegebener Trefferwahrscheinlichkeit p liegt die (zufällige) relative Trefferhäufigkeit $H = X/n$ mit ca. $95\ \%$ Sicherheit im Prognoseintervall

$$\left[p-2\frac{\sqrt{p(1-p)}}{\sqrt{n}};\ p+2\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \subseteq \left[p-\frac{1}{\sqrt{n}};\ p+\frac{1}{\sqrt{n}}\right]. \qquad (1)^6$$

Seine Länge schrumpft umgekehrt proportional mit der Wurzel aus dem Versuchsumfang. Sie halbiert sich, wenn man den Versuchsumfang vervierfacht.

Abb. 2 und **3** wurden mit der Datei *eins-durch-wurzel-n.ggb* erzeugt. Sie veranschaulichen diese Aussage durch sukzessiv eingeblendete relative Häufigkeitstrajektorien, die bei $n = 25, 100$ und 400 in **Abb. 2** „vertikal streuende" Spurpunkte hinterlassen, die man in **Abb. 3** auch durch horizontal liegende Boxplots veranschaulicht.

- An jeder festen Stelle liegen etwa 95 % dieser Spurpunkte im Wurzeltrichter, der durch die Graphen von $f_{\mp}$ mit $f_{\mp}(p) = p \mp 2\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ begrenzt wird.
- Der Anteil der Trajektorien, die bis zur Stelle n komplett im Wurzeltrichter liegen, ist aber deutlich kleiner als $95\ \%$, umso kleiner, je größer n ist.
- Nur in ca. $2/3$ aller Fälle rutscht die relative Häufigkeit näher an p heran, wenn man den Versuchsumfang vervierfacht. Der mittlere Abstand zu p verringert sich stets.
- Wenn man die Streuung der relativen Häufigkeiten wie in **Abb. 3** ab Klasse 7 durch Boxplots visualisiert, erkennt man, dass die Quartilabstände sich ungefähr halbieren, wenn man den Versuchsumfang vervierfacht. Und schon Achtklässler entdecken den „Hauptsatz der beurteilenden Statistik": „Viermal so viel, doppelt genau" auch ohne Kenntnis von Wurzeln und Standardabweichung. Er liefert die Begründung dafür, dass man Wahrscheinlichkeitsangaben umso mehr Vertrauen schenken darf, je mehr Experimente dahinterstecken.

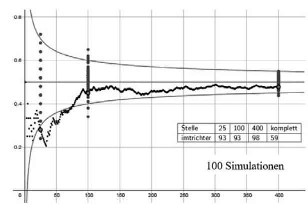


Abb. 2 Wurzeltrichter: *p* fest, *n* variabel, hier *n* = 400

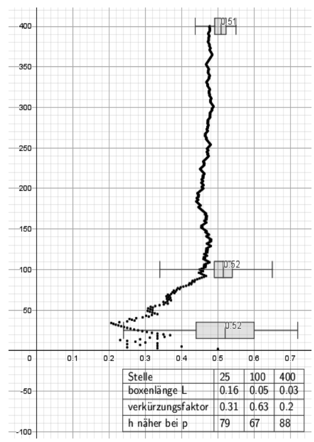


Abb. 3

In der Sekundarstufe I hat das $1/\sqrt{n}$ -Gesetz den Status eines Naturgesetzes, vgl. RIEMER [1991]. Dabei genügt die Untersuchung von $p = 1/2$ bzw. die Abschätzung in (1). In der Sekundarstufe II begründet man es mit der 2σ-Regel der Binomialverteilung wie folgt:

Bei der Binomialverteilung mit $\mu = n \cdot p$, $\sigma = \sqrt{n \cdot p \cdot (1-p)}$ liegt die Trefferanzahl X mit ca. $95\ \%$ Sicherheit im 2σ-Prognoseintervall

$$\left[np-2\sqrt{np(1-p)};\ np+2\sqrt{np(1-p)}\right]. \qquad (2)$$

6 wegen $p(1-p) \leq 1/4$

Weil die Länge dieses Intervalls proportional zu $\sqrt{n}$ wächst, spricht man bei (2) auch vom $\sqrt{n}$ -Gesetz der absoluten Häufigkeiten. Wenn man durch n dividiert, erhält man hieraus das $1/\sqrt{n}$ -Gesetz (1) der relativen Häufigkeiten.

Falls die relative Häufigkeit h nicht in diesem Intervall liegt, dann nennt man die Abweichung von p *statistisch signifikant*. Wenn man sich bei p nicht sehr sicher ist – und eigentlich auch nur dann – beginnt man zu *bezweifeln, dass p ein gutes Modell darstellt*. Man begibt sich ggf. auf die Suche nach besseren Modellen oder arbeitet mit p in dem Bewusstsein weiter, dass man vielleicht kein optimales Modell erwischt hat. Diese intuitiv eingängige Idee nennen wir das *Konzept des Bezweifelns*. Es handelt sich dabei um eine auf das Wesentliche reduzierte Variante des zweiseitigen Signifikanztests auf dem 5 %-Niveau.

2.2 Wie unterrichtet man Prognoseintervalle?

Nach einem Blick auf empfehlenswerte digitale Werkzeuge skizzieren wir (über ausformulierte Arbeitsaufträge) bewährte Unterrichtsstunden und Aufgabenstellungen.

2.2.1 Digitale Werkzeuge

Vorab zeigen wir in **Abb. 4** bis **Abb. 6**, mit welchen digitalen Werkzeugen wir neben der oben erwähnten Datei *eins-durch-wurzel-n.ggb* aus **Abb. 2** und **Abb. 3** ausgezeichnete Erfahrungen gemacht haben. Dabei kommt es uns nicht in erster Linie auf die Werkzeuge an, die sind schnell programmiert, sondern auf präzise formulierte und herausfordernde Problemstellungen, die wir Lernenden *schenken*! Dabei ist für uns der Dreischritt „**Spekulieren – Experimentieren – Reflektieren**“ handlungsleitend. Wenn experimentiert wird, dann *zuerst händisch*, damit Schülerinnen und Schüler *verstehen*, was *anschließend digital* abläuft: **„Verstehen“ hat viel zu tun mit Verlangsamen bis hin zum „Stehenbleiben“.**

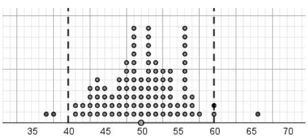


Abb. 4 Simulation einer Stichprobenverteilung, absolute Häufigkeiten

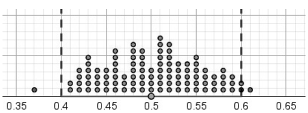


Abb. 5 Simulation einer Stichprobenverteilung; relative Häufigkeiten

Abb. 4 wurde mit der GeoGebra-Datei *galtonabsolut.ggb* erzeugt. Hier werden Pseudozufallszahlen von einer Binomialverteilung mit n Versuchen und der Trefferwahrscheinlichkeit p erzeugt. Es ist vorteilhaft, sich einen konkreten Zufallsversuch vorzustellen, z. B. ein Glücksrad mit Trefferwahrscheinlichkeit p: Es wird n-mal gedreht und die Anzahl der Treffer wird als Punkt über der Trefferanzahl auf der Rechtsachse visualisiert. Dies wird w-mal wiederholt (in diesem Fall 100-mal). Jedes n-malige Drehen liefert eine Trefferanzahl, die als weiterer Punkt dargestellt wird. Durch Wiederholung des Experiments entsteht ein „glockenförmiges“ Säulendiagramm (eine Stichprobenverteilung) als Realisierung der Binomialverteilung $B(n,p)$. Das 2σ-Prognoseintervall kann zusätzlich eingeblendet werden.

Abb. 5 wurde mit der GeoGebra-Datei *galtonrelativ.ggb* erzeugt. Wie in **Abb. 4**, nur dass die Säulendiagramme über den Stellen der relativen Häufigkeiten entstehen.

Wenn n erhöht wird, erkennt man, dass die Säulendiagramme der absoluten Häufigkeiten (**Abb. 4**) immer breiter werden und „nach rechts weglaufen“. Die Säulendiagramme der relativen Häufigkeiten (**Abb. 5**) konzentrieren sich dagegen in der Nähe von p.

Die Datei *prognoseintervall-tool.ggb* aus **Abb. 6** dient als Rechenwerkzeug: In Abhängigkeit von der „verschiebbaren“ Wahrscheinlichkeit p auf der Rechtsachse werden die Prognoseintervalle der relativen Häufigkeiten als „lotrechte“ Strecken zwischen den Funktionsgraphen von $f_{\mp}$ mit $f_{\mp}(p) = p \mp 2\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ in einem Ellipsendiagramm veranschaulicht, das die Brücke zu Konfidenzintervallen vorbereitet.

Da in einigen Bundesländern die LAPLACE-Bedingung einen gewissen Stellenwert besitzt, wird die Ellipse über den Stellen dünner dargestellt, an denen die LAPLACE-Bedingung $np(1-p) > 9$ nicht erfüllt ist.

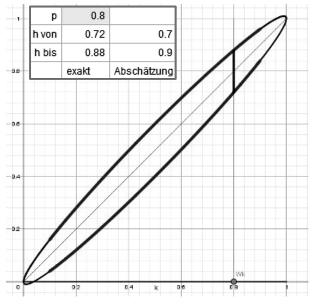


Abb. 6 Ellipse $n = 100$ fest, p variabel

2.2.2 Prognoseintervalle: kommentierte Arbeitsblätter und Aufgaben

2.2.2.1 Prognoseintervalle qualitativ (Arbeitsblatt für Klasse 8 – 10)

Wahrscheinlichkeiten schauen in die Zukunft! Sie drücken aus, welche relative Häufigkeit man für ein Ereignis erwartet. Damit ist nicht gemeint, dass die relative Häufigkeit genau diesen Wert annehmen wird (was sehr unwahrscheinlich, bis unmöglich sein kann), sondern dass die relative Häufigkeit in der Nähe dieses Wertes liegt, also „nicht allzu sehr“ davon abweicht. Was „nicht allzu sehr“ bedeutet, das soll in diesem Arbeitsauftrag – in Abhängigkeit vom Versuchsumfang n – erforscht werden.

Spekulieren: ein Gedankenexperiment

a) Wirf *in Gedanken* 25-mal eine faire Münze. Wappen gilt als Treffer. Notiere vier dir *plausibel erscheinende* Trefferzahlen zwischen 0 und 25, analog auch für 100 und für 400 Münzwürfe.

b) Speichert eure geschätzten Trefferzahlen in den Spalten B – D der Datei *schwankung-schätzen.ggb*. Die zugehörigen relativen Häufigkeiten werden in den Spalten E – F berechnet und als Punktwolken über den Stellen 25, 100 und 400 veranschaulicht.

c) Diskutiert mithilfe der grafischen Darstellungen eure Intuitionen. Vergleicht mit dem Diagramm (**Abb. 7**) einer Klasse 9. Haltet hierzu Aussagen oder Fragen in Stichworten fest.

	A	B	C	D	E	F	G
2		25	100	400	25	100	400
3	bernd 1	15	17	75	0.6	0.17	0.19
4	bernd 2	13	31	200	0.52	0.31	0.5
5	bernd 3	19	68	150	0.76	0.68	0.38
6	bernd 4	9	48	260	0.36	0.48	0.65
7	denise 1	14	56	220	0.56	0.56	0.55
8	denise 2	8	39	239	0.32	0.39	0.6
9	denise 3	9	41	179	0.36	0.41	0.45
10	denise 4	13	63	211	0.52	0.63	0.53

Abb. 7a Schätzungen einer Klasse 9

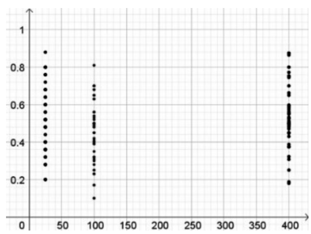


Abb. 7b Zugehörige Grafik

Experimentieren (erst händisch, dann digital)

a) Führt 100 Münzwürfe durch. Jeder diktiert seine vier relativen Häufigkeiten zu 25 und eine zu 100 Münzwürfen (Vierergruppen zusammen auch eine zu 400 Münzwürfen) in eine Kopie von *schwankung-schaetzen.ggb*. Vergleicht mit euren Schätzungen aus a) – c).
Lasst den Computer 100 Serien mit je 25, 100 und 400 Münzwürfen ausführen.

b) Stellt die relativen Trefferhäufigkeiten h_{25}, h_{100}, h_{400} nach 25, 100 und 400 dar wie in **Abb. 7b.** Vergleicht mit euren Schätzungen aus a) – c).

Das Spekulieren vor dem Experimentieren sorgt hier für einen Spannungsbogen. In der Regel unterstellen Schülerinnen und Schüler neben der selbstverständlichen Proportionalität zwischen Versuchsumfang und erwarteter Trefferzahl „primärintuitiv“ nämlich auch eine Proportionalität zur Größe der Zufallsschwankungen, die dann tendenziell konstante Zufallsschwankungen bei den relativen Häufigkeiten impliziert (vgl. **Abb. 7b**).

Damit entsteht ein kognitiver Konflikt zur Intuition „je mehr, desto genauer“. Die händischen und digitalen Simulationen lösen diesen Widerspruch auf. Sie sichern das empirische Gesetz der großen Zahlen qualitativ ab: „Mit steigender Versuchszahl nehmen die Zufallsschwankungen der relativen Häufigkeiten um die Wahrscheinlichkeit tendenziell ab.“ *Der Nachsatz, dass man deswegen der Schätzung einer Wahrscheinlichkeit umso mehr Vertrauen entgegenbringt, je größer der zugrunde liegende Versuchsumfang war, bereitet die Idee der Konfidenzintervalle schon in der Sekundarstufe I vor.*

2.2.2.2 Prognoseintervalle quantitativ

Die folgenden Arbeitsaufträge zeigen, wie Schülerinnen und Schüler z. B. im Anschluss an 2.2.2.1 die Simulationen in **Abb. 2** und **3** nutzen können, um das Gesetz der großen Zahlen nicht nur qualitativ, sondern auch quantitativ zu erforschen. Dabei wird auch hier wieder auf verlangsamendes, arbeitsteiliges, die Kommunikation untereinander förderndes Vorgehen Wert gelegt. Da in Klasse 7/8 noch keine Wurzeln (wie in **Abb. 2**) verfügbar sind, nutzt man die Boxenlänge (**Abb. 3**) als Streuungsmaß.

Abnahme von Zufallsschwankungen – Boxplots (Klasse 7 – 8)

Wenn man eine Münze mit Trefferwahrscheinlichkeit p insgesamt 400-mal wirft, spricht man von einer BERNOULLIkette[7] der Länge 400. Der „gezackte“ Häufigkeitsgraph der Zuordnung $n \to h_n$ (**Abb. 2**) veranschaulicht, wie sich die relative Häufigkeit h_n bei einer solchen Kette mit $p = 0{,}5$ entwickelt. Dabei wurden die relativen Häufigkeiten nach 25 bzw. 100 bzw. 400 Würfen (h_{25} bzw. h_{100} bzw. h_{400}) markiert. Die Markierungen sind Spuren, die solche Ketten nach 25 bzw. 100 bzw. 400 Würfen hinterlassen.

p=0.5	n	L_{25}	L_{100}	L_{400}	L_{100} / L_{25}	L_{400} / L_{100}	$\|h_{100}-p\|<\|h_{25}-p\|$	$\|h_{400}-p\|<\|h_{100}-p\|$
Walter	100	0.2			0.38	0.52	72%	74%
Doris								
Mittelwerte								

Abb. 8 Jeder trägt seine Simulationsergebnisse in eine gemeinsame Tabelle ein.

a) Formuliert eine Aussage, die ihr **Abb. 2** entnehmt.
b) Erklärt, wie **Abb. 3** mit **Abb. 2** zusammenhängt. Spekuliert über quantitative Aussagen, die man den abgebildeten Boxplots entnehmen kann. Tipp: Die Länge (L_{25}, L_{100}, L_{400}) der Boxen (der „Quartilabstand“) ist jeweils eingeblendet.
c) Prüft eure Vermutungen. Simuliert dazu viele BERNOULLIketten. Füllt die ersten fünf Spalten in **Abb. 8** gemeinsam aus, bildet Mittelwerte und haltet die Ergebnisse in Stichworten fest.
d) In **Abb. 3** ist zu erkennen, wie oft die relative Häufigkeit h durch Vervierfachung des Versuchsumfanges (von 25 auf 100 und von 100 auf 400) näher an die Wahrscheinlichkeit p heranrutscht – formal, wie oft gilt $|h_{100} - p| < |h_{25} - p|$ und $|h_{400} - p| < |h_{100} - p|$.

7 Natürlich kann man BERNOULLIketten schlicht als „Münzwurfserien“ bezeichnen und sich auf $p = 1/2$ beschränken.

Sammelt auch diese Ergebnisse in den beiden letzten Tabellenspalten und formuliert ein Forschungsergebnis, das wie folgt beginnen könnte: „Wenn man bei einer BERNOULLI-kette den Versuchsumfang vervierfacht, dann …“

Das $1/\sqrt{n}$ -Gesetz mit Wurzeltrichter (Klasse 9/10)

Abb. 2 zeigt die Entwicklung der relativen Trefferhäufigkeit h_n in einer BERNOULLIkette.

Es werden die Wurzelfunktionen $f_{\mp}$ mit $f_{\mp}(p) = p \mp 2\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ eingeblendet.

Sie begrenzen einen „Wurzeltrichter“ um die Wahrscheinlichkeit *p*.

a) Formuliere eine Vermutung über den Zusammenhang zwischen den drei Häufigkeitsgraphen und dem Wurzeltrichter.

b) Simuliere 100 BERNOULLIketten der Länge 400. Zähle mithilfe des Programms, wie oft die relative Häufigkeit h_n an den Stellen $n = 25,\ 100,\ 400$ innerhalb des Wurzeltrichters liegt und zusätzlich, wie oft der komplette Häufigkeitsgraph (für *alle* Werte von *n*) innerhalb des Wurzeltrichters verläuft.

		innerhalb des Wurzeltrichters			
p=0.5	n	r_{25}	r_{100}	r_{400}	komplett
Kim	100				
Matthias					
Summe					
Mittel					

Abb. 9 Tabelle zum Eintragen der Ergebnisse

c) Übertragt die Simulationsergebnisse in eine gemeinsame Tabelle (**Abb. 9**). Wertet die Daten durch Ausfüllen der letzten beiden Tabellenzeilen aus. Vergleicht mit euren Vermutungen aus a) und formuliert ein Forschungsergebnis: „Ein Schätzwert der Wahrscheinlichkeit dafür, dass … im Wurzeltrichter liegt …“

2.2.2.3 Die Wette gilt! – Ein alternativer Einstieg in Prognoseintervalle

Auch die folgende Einstiegsvariante ist prinzipiell für die Sekundarstufe I geeignet. Durch die direkte Anschlussfähigkeit an das Thema Binomialverteilung und 2σ-Regel ist eine Verortung in der Sekundarstufe II aber naheliegender. Die 2σ-Regel wird anschließend im Unterricht mit der Binomialverteilung kontrolliert und führt dann zur Begründung des $1/\sqrt{n}$ -Gesetzes und über das Konzept des Bezweifelns in die beurteilende Statistik.

Ähnlich wie 2.2.2.1 beginnt man mit Spekulationen über die Größe von Zufallsschwankungen, die man über eine Wettsituation und den Dreischritt *Spekulieren – Experimentieren – Reflektieren* im Motivationshorizont der Lernenden verankert.

Die Wette gilt! (Klassenstufe 11)

a) Angebot: „Ihr seid 30 Schüler. Ihr bekommt mein Fahrrad, wenn eine(r) von euch bei 100 Münzwürfen weniger als 30- oder mehr als 70-mal „Wappen“ erhält. Wenn das keiner von euch schafft, ladet ihr mich zu einem Kaffee ein. Die Wette gilt! Diskutiert nach Bauchgefühl, was ihr von diesem Angebot haltet. Vergleicht mit folgenden Angeboten:
   - Fahrrad bei weniger als dreimal Zahl oder weniger als dreimal Wappen bei 10 Münzwürfen,
   - Fahrrad bei weniger als 120-mal Zahl oder weniger als 120-mal Wappen bei 400 Münzwürfen.

   Führt die Experimente mit den 10 Münzwürfen händisch durch und fasst je 10 Ergebnisse zu einem 100er-Experiment zusammen. Diskutiert erneut und gebt eine erste Bewertung ab.“

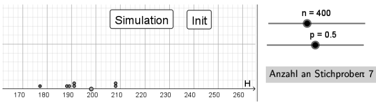


Abb. 10 Sieben absolute Trefferhäufigkeiten bei 400 Würfen einer LAPLACE-Münze

Nina hat mit *galton-absolut.ggb* die **Abb. 10** erzeugt. „Hinter den 7 Punkten stecken aber verdammt viele Münzwürfe“, meint sie. Deutet die Punkte und erläutert Ninas Aussage.

b) Untersucht das Angebot mit *galton-absolut.ggb* arbeitsteilig.
Haltet die Forschungsergebnisse in Stichworten fest und bewertet eure Schätzungen aus a) erneut.

2.2.2.4 Top-Down-Einstieg: Die Faustregel vorgeben – den Sinn erschließen

Der folgende Arbeitsauftrag zeigt, dass auch ein Top-Down-Vorgehen effektiv und sinnstiftend sein kann. Man gibt die $1/\sqrt{n}$ -Faustregel vor und erschließt deren Bedeutung unter Rückgriff auf konkrete Beispiele durch einen Vergleich mit dem „Bauchgefühl“ [vgl. BRANDT u. a. 2018, S. 122].

Das Konzept des Bezweifelns (händisch, Klasse 12)

Fußlage F

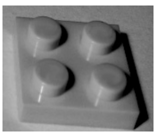


Abb. 11 Flacher Legostein

Ein flacher Legostein (**Abb. 11**) kann auf dem Fuß oder auf dem Kopf landen.
MAX: „Fußlage ist wahrscheinlicher als Kopf, weil Fuß eine größere Auflagefläche hat.“
HANNAH: „Kopflage ist wahrscheinlicher, weil die Noppenseite schwer ist und eher unten liegen wird.“

a) Würfle so lange, bis du nach DEINEM Bauchgefühl sicher bezweifeln kannst, dass gilt $p(\text{Fuß}) = 1/2$.
Notiere, ob deiner Meinung nach eher MAX recht hat oder eher HANNAH!
b) Da jeder ein anderes Bauchgefühl hat, haben Statistiker folgende Faustregel aufgestellt:

> **Faustregel:** Bezweifle eine Wahrscheinlichkeitsangabe p, wenn bei einer Stichprobe vom Umfang n die relative Häufigkeit h um mehr als $2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{1}{\sqrt{n}}$ von p abweicht.
> Man bezeichnet die Abweichung dann als statistisch **signifikant**.

c) Überprüfe deine Bauchentscheidung mit der Faustregel und beurteile, wer schneller Zweifel signalisiert: dein Bauch oder die Faustregel.

2.2.2.5 Erhellende Aufgaben

Das Konzept des Bezweifelns und der Stichprobenumfang (Klasse 12)

Wie sicher erkennt man mithilfe der Faustregel gezinkte Münzen?

a) Simuliere den $n = 25$-fachen Wurf einer *gezinkten Münze* mit $p\,(\text{Wappen}) = 0{,}7$ und bestimme anschließend die relative Häufigkeit h (Wappen).
b) Prüfe, ob man nach der Faustregel die „Fairness“ der Münze: $p\,(\text{Wappen}) = 0{,}5$ schon nach 25 Münzwürfen bezweifeln kann.
c) Wiederhole die Simulation mehrfach und schätze die Wahrscheinlichkeit dafür, dass man gemäß Faustregel bei einer mit $p = 0{,}7$ gezinkten Münze nach 25 Würfen $p = 0{,}5$ die „Fairness“ der Münze bezweifeln sollte.
d) Führt die Untersuchung c) arbeitsteilig durch mit $n = 100$ und $n = 400$.

e) Sichert die Simulationsergebnisse rechnerisch mithilfe der Binomialverteilung ab.
f) Fasst die Erkenntnisse aus c) und d) in Worten zusammen.

Abb. 12 vermittelt einen Eindruck, mit welchen Lösungen man bei den Aufgabenteilen c), d) und f) in einem Grundkurs rechnen kann.

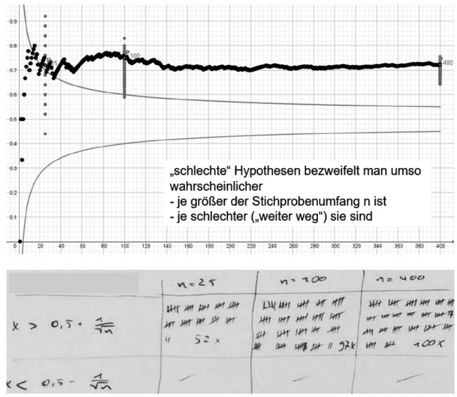


Abb. 12 Beispiele von Simulationen zu c) – f); Lösung eines Schülers

Das Eichzeichen (Klasse 12)

Das Eichzeichen „e“ (**Abb. 13**) garantiert, dass die Gewichtsangabe 1,8 g (netto) im Mittel eingehalten wird. Weil Produzenten nichts verschenken, liegt die Annahme nahe, dass die Tütchen mit Wahrscheinlichkeit 50 % über- und mit 50 % untergewichtig sind.

a) Bestimme mit dieser Annahme für den Stichprobenumfang $n = 144$ das Prognoseintervall für die relative Häufigkeit $h(+)$ übergewichtiger Tütchen.
b) NICO: „Ich hatte 56 % übergewichtige Tütchen. Das ist eine signifikante Abweichung nach oben und ich darf bezweifeln, dass ein Tütchen nur mit Wahrscheinlichkeit 50 % übergewichtig ist.“
Überprüfe und kommentiere NICOs Aussage.

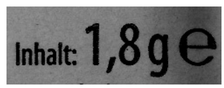


Abb. 13 Präzisionswaagen zeigen nie genau 1,800 g, stets etwas weniger oder etwas mehr.

Signifikanz – Relevanz (Klasse 12)

Der Unterschied zwischen den beiden Wahrscheinlichkeiten $p_1 = 0{,}50$ und $p_2 = 0{,}51$ ist in der Praxis irrelevant. Begründe mithilfe des $1/\sqrt{n}$ -Gesetzes,

a) wenn man hinreichend viele Versuche macht, ist es sehr wahrscheinlich (95 %), dass man $p_1 = 0{,}50$ bezweifeln muss, wenn tatsächlich $p_2 = 0{,}51$ gilt.
b) Berechne, wie viele Versuche man machen müsste.
c) HANNAH: „Wenn ich sehr viele Versuche machen darf, dann kann ich jede Hypothese mit hoher Wahrscheinlichkeit (95 %) bezweifeln.“ Begründe HANNAHs Aussage.

3 Konfidenzintervalle

3.1 Worum geht es?

Man definiert das Konfidenzintervall I zu einer relativen Häufigkeit h als „Ansammlung“ aller Wahrscheinlichkeiten p, die nach Beobachtung von h nicht bezweifelt zu werden brauchen. *Das Konfidenzintervall enthält also alle diejenigen Wahrscheinlichkeiten p, in deren Prognoseintervall die beobachtete relative Häufigkeit h liegt.* Da man Wahrscheinlichkeiten, die man nicht bezweifeln muss, ein gewisses Vertrauen entgegenbringt (Lat.: confidere = vertrauen), erklärt sich der Name Konfidenzintervall von selbst.

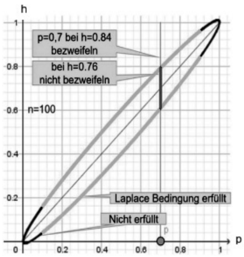
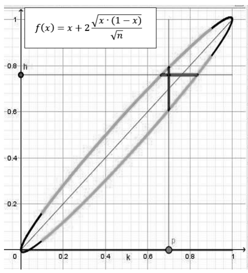
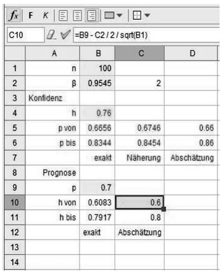


Abb. 14 Prognoseellipse

Am nebenstehenden Ellipsendiagramm erkennt man: Eine fragliche Wahrscheinlichkeit von $p = 0{,}7$ muss bei Beobachtung der relativen Häufigkeit

- $h = \frac{84}{100} = 0{,}84$ bezweifelt werden, denn $h = 0{,}84$ liegt außerhalb des markierten Prognoseintervalls zu $p = 0{,}7$. Daher gehört $p = 0{,}7$ nicht in das Konfidenzintervall zu $h = 0{,}84$.
- $h = \frac{76}{100} = 0{,}76$ nicht bezweifelt werden, denn $h = 0{,}76$ liegt im (markierten) Prognoseintervall zu $p = 0{,}7$. Daher gehört $p = 0{,}7$ in das Konfidenzintervall zu $h = 0{,}76$.





C10 =B9 - C2 / 2 / sqrt(B1)

	A	B	C	D
1	n	100		
2	β	0.9545	2	
3	Konfidenz			
4	h	0.76		
5	p von	0.6656	0.6746	0.66
6	p bis	0.8344	0.8454	0.86
7		exakt	Näherung	Abschätzung
8	Prognose			
9	p	0.7		
10	h von	0.6083	0.6	
11	h bis	0.7917	0.8	
12		exakt	Abschätzung	
13				
14				

Abb. 15 Entdeckungen mit der GeoGebra-Datei *prognose-konfidenzintervalle-werkzeug.ggb*

Man sieht in **Abb. 15** durch Verschieben von p auf der Rechtsachse auch, dass etwas kleinere und etliche größere Wahrscheinlichkeiten als $p = 0{,}7$ im Konfidenzintervall zu $h = 0{,}76$ liegen.

Die Frage, welche Wahrscheinlichkeiten genau ins Konfidenzintervall zu $h = 0{,}76$ gehören, wird grafisch beantwortet durch die Stellen $\left[\approx 0{,}66; \approx 0{,}83\right]$ auf der Rechts-(p-)Achse, in denen die Horizontale zu $h = 0{,}76$ die „Ränder der Ellipse" schneidet.

Rechnerisch bestimmt man sie durch (numerisches) Lösen der Gleichungen:

$$f_+(p) = 0{,}76 \text{, also } h = p + 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \text{ zu } p_1 \approx 0{,}6656 \text{ (linke Grenze) bzw.} \qquad (*)$$

$$f_-(p) = 0{,}76 \text{, also } h = p - 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \text{ zu } p_2 \approx 0{,}8344 \text{ (rechte Grenze).} \qquad (**)$$

Man kann die Gleichungen auch umformen und erhält nach Lösung einer (länglichen) quadratischen Gleichung für das Konfidenzintervall

$$I = \left[\frac{h + \frac{2}{n} - 2\sqrt{\frac{1}{n^2} + \frac{h(1-h)}{n}}}{\frac{4}{n} + 1}; \frac{h + \frac{2}{n} + 2\sqrt{\frac{1}{n^2} + \frac{h(1-h)}{n}}}{\frac{4}{n} + 1}\right] \text{ „exakt" (WILSON-Intervall).} \qquad (***)$$

Für großes n ergibt sich hieraus die von Taschenrechnern angebotene Näherung

$$I \approx \underbrace{\left[h - 2\frac{\sqrt{h(1-h)}}{\sqrt{n}}; h + 2\frac{\sqrt{h(1-h)}}{\sqrt{n}}\right]}_{\text{„Näherung" (WALD-Intervall)}} \subseteq \underbrace{\left[h - \frac{1}{\sqrt{n}}; h + \frac{1}{\sqrt{n}}\right]}_{\text{„Überschlag"}}, \qquad (****)$$

wobei sich das rechts stehende Intervall als Faustformel für Überschläge anbietet.

Im Unterricht wird man, nachdem die Zusammenhänge verstanden und gesichert sind, Routinerechnungen an ein Tabellenkalkulationsblatt wie in **Abb. 14** delegieren. Dort stehen die „exakten“[8] (WILSON-), gerundeten (WALD-) und überschlagenen Konfidenzintervalle im Bereich B5 : D6. (Ergänzend findet man die „exakten“ und überschlagenen Prognoseintervalle im Bereich B10 : C11.)

Der Unterricht ist dann erfolgreich, wenn folgende Aussagen zum Kopfwissen werden:

- Prognoseintervalle enthalten relative Häufigkeiten. Sie sind symmetrisch zur Wahrscheinlichkeit p.
- Im Prognoseintervall zu einer bekannten Wahrscheinlichkeit p liegt die relative Häufigkeit h mit der Wahrscheinlichkeit $\gamma = 95\ \%$. Man nennt $\gamma = 95\ \%$ **Sicherheitswahrscheinlichkeit**.
- Konfidenzintervalle enthalten Wahrscheinlichkeiten. WILSON-Konfidenzintervalle sind nur für $h = 0{,}5$ symmetrisch zur relativen Häufigkeit h. Die WALD-Konfidenzintervalle sind stets symmetrisch zu h.
- Das (zufallsabhängige) Konfidenzintervall zu einer beobachteten relativen Häufigkeit h überdeckt die unbekannte (feste) Wahrscheinlichkeit p mit der Wahrscheinlichkeit $\gamma = 95\ \%$. Man nennt $\gamma = 95\ \%$ das Konfidenzniveau (oder Vertrauenswahrscheinlichkeit).
- Prognoseintervalle halbieren sich bei Vervierfachung des Versuchsumfanges. WILSON-Konfidenzintervalle halbieren sich bei Vervierfachung des Versuchsumfanges annähernd, WALD-Konfidenzintervalle genau.

3.2 Wie unterrichtet man Konfidenzintervalle?

Wir beantworten diese Frage wieder durch das Angebot kommentierter Aufgabenstellungen: 3.2.1 zeigt, wie man den flachen Legostein (als Prototyp einer unbekannt, aber „normiert (!) gezinkten Münze“) auch zur Einführung des Begriffs „Konfidenzintervall“ nutzt. Beim Experimentieren, Konstruieren und Auswerten erleben Schülerinnen und Schüler die Zufallsabhängigkeit von Konfidenzintervallen, die sich beim Aufgabenlösen ansonsten schnell hinter Rechnungen versteckt. Das ist für die Begriffsentwicklung und sachgerechte Interpretation von unschätzbarem Wert. 3.2.2 lädt ein zur Entdeckung von Konfidenzintervallen über eine digitale Lernumgebung. 3.2.3 unterstützt die sachgerechte Interpretation der Information, die Konfidenzintervalle bieten, mithilfe einer Simulation. Empfehlenswerte Aufgaben runden die Darstellung ab. Weitere Experimente und Beispiele findet man in [RIEMER 2019].

3.2.1 Intuitive Begriffsbildung

Spekulieren

Durch die Aufforderung „Notiere ein Intervall, in dem du (vor jeglichem Experiment) beim flachen Legostein p (Fußlage) vermutest“ wird der Begriff des Konfidenzintervalls – als *Intervall von Wahrscheinlichkeiten* – intuitiv gebildet und vom bereits gesicherten Begriff des Prognoseintervalls als *Intervall relativer Häufigkeiten* abgegrenzt. Unterschiedliche Einschätzungen (**Abb. 16**) sorgen auch hier für Spannung.

"Bauch" - Konfidenzintervall

min	max	unter 0.5	linkslastig	rechtslastig	über 0.5
		5	8	4	6
0,4	0,52		1		
0,55	0,7				1
0,55	0,65				1
0,35	0,45	1			
0,4	0,55		1		
0,4	0,5		1		
0,37	0,47	1			
0,5	0,6			1	
0,45	0,6			1	
0,6	0,7				1

Abb. 16 Intervallschätzungen von *p* (Fußlage) beim flachen Legostein

[8] „Exakt“ ist nur im Sinne der exakten Lösung der Gleichungen (*) und (**) zu verstehen, die aber selbst nur auf den Sigmaregeln, also nur auf Näherungsbeziehungen, beruhen.

Die Intervalle bezeichnen wir unter bewusster Verwendung einer schülernahen Arbeitssprachebene als „Bauch-Konfidenzintervalle“, weil sie nur durch eine Intuition, ein vages „Bauchgefühl“, begründet sind.

Experimentieren

- Jede(r) „würfelt“ mit seinem flachen Lego 100-mal und prüft anschließend mit Prognoseintervallen, ob die Versuchsergebnisse Anlass liefern, $p(\mathrm{F}) = 0{,}1$, $p(\mathrm{F}) = 0{,}2$, $\ldots p(\mathrm{F}) = 0{,}8$, $p(\mathrm{F}) = 0{,}9$ zu bezweifeln.
- Bei ANTONIA ($h = 35/100$) müssen nur 0,3 und 0,4 nicht, bei MIRJAM ($h = 47/100$) nur 0,4 und 0,5 nicht bezweifelt werden (s. **Abb. 17** und **Abb. 18**).

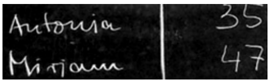


Abb. 17 Zwei Wurfergebnisse

p	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
bis	0.000	0.160	0.280	0.392	0.498	0.600	0.698	0.792	0.880	0.960	1.000
von	0.000	0.040	0.120	0.208	0.302	0.400	0.502	0.608	0.720	0.840	1.000
100				A	A M	M					

Abb. 18 Berechnen und interpretieren

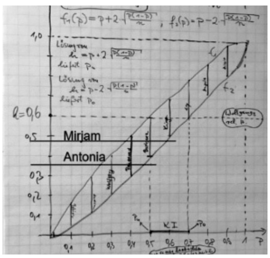


Abb. 19 Händisch erstellte Konfidenzellipse

- Nach dieser Aufgabenstellung beginnen die Schülerinnen und Schüler von selbst, also ohne dass sie dazu aufgefordert werden müssten, das jeweilige Intervall der nicht zu bezweifelnden Wahrscheinlichkeiten durch numerisches Probieren einzugrenzen (wie bei der Intervallschachtelung zu $\sqrt{2}$). Was dabei herauskommt, ist dann das nicht mehr vom Bauchgefühl, sondern „nur noch“ vom Versuchsergebnis abhängige (!) „objektive Konfidenzintervall“. ANTONIA erhält durch Probieren $[0{,}262; 0{,}449]$ und MIRJAM $[0{,}373; 0{,}569]$. Unterschiedliche Erfahrungen führen halt zu unterschiedlichen Erwartungen und unterschiedlichen Konfidenzintervallen.
- Das händische Probieren ist ein Schlüssel zum *selbstständigen* Entdecken der Gleichungen (*) und (**), also zur Bestimmung der „objektiven“ Konfidenzintervalle. Gestützt durch die „emotionale Eingebundenheit“ und den Vergleich der „objektiven“ mit den Bauch-Konfidenzintervallen erschließt sich die Bedeutung der „neuen“ Intervalle gleichsam nebenbei.
- Dabei erweist sich das arbeitsteilige Berechnen und Einzeichnen der Prognoseintervalle in ein gemeinsames Plakat (**Abb. 19**) gerade wegen seiner händischen Unzulänglichkeit als hilfreich für sinnstiftende Diskussionen. Personale Verknüpfungen wie „Das Konfidenzintervall zu WOLFGANGS relativer Häufigkeit 60/100 reicht von BARBARA bis EVY“ bleiben hängen und signalisieren: Mathe hat mit mir zu tun. Ein solches Plakat schlägt die Brücke zu dem Rechenwerkzeug aus **Abb. 14**, das dafür sorgt, dass das Nachdenken nicht im Rechnen erstickt.

3.2.2 Konfidenzintervalle freirubbeln

Die Datei *konfidenzintervall-freirubbeln.ggb*[9] (**Abb. 19**) dient der Begriffsentwicklung und zum Entdecken des Sinns von Konfidenzintervallen. Dies geschieht nach dem Blackbox-Whitebox-Prinzip der CAS-Didaktik, bei der ein vorgefertigtes Produkt (Blackbox) han-

[9] Dies ist die digitale Variante des in Niedersachsen populären „VEHLINGschen Zollstocks“, eines Partnerspiels, bei dem die Wahrscheinlichkeit durch eine auf einem Zahlenstrahl „verschiebbare“ Person dargestellt wird, die einen – das Prognoseintervall veranschaulichenden – Zollstock trägt. Eine zweite Person steht dabei fest auf der relativen Häufigkeit.

delnd mit Sinn gefüllt (zur Whitebox) wird: Eine beobachtete relative Häufigkeit wird (als Punkt im Intervall $[0;1]$) markiert. Eine Wahrscheinlichkeit p ist (samt zugehörigem Prognoseintervall) auf der Rechtsachse „probeweise" verschiebbar. Die probierte Wahrscheinlichkeit wird genau dann durch einen eingeblendeten „Zwillingspunkt" dupliziert, wenn die „beobachtete" relative Häufigkeit h im Prognoseintervall der Wahrscheinlichkeit p liegt, wenn also das probierte p zum beobachteten h „passt". Durch Verschieben der Wahrscheinlichkeit wird im Spurmodus das Konfidenzintervall gleichsam „freigerubbelt". Das verlangsamende „Punkt-für-Punkt"-Freirubbeln regt die Begriffsbildung „Konfidenzintervall" an und führt letztendlich zu einer begrifflich sauberen Definition. Die gewollte numerische Ungenauigkeit wird als Motivationsanker für die Entwicklung einer Formel genutzt. Eine Aufgabenstellung sieht mit **Abb. 20** wie folgt aus:

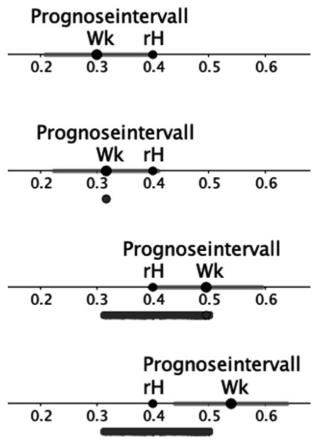


Abb. 20 Kl freirubbeln

Beobachtet wurde bei 100 Versuchen die relative Häufigkeit $h = 0{,}4$.

a) HANNAH: „Das Prognoseintervall um p belegt: Ich darf bezweifeln, dass diese relative Häufigkeit von $p = 0{,}3$ stammt." Überprüfe durch Nachrechnen.

b) NIKO: „Wenn ich p auf 0,31 erhöhe, signalisiert der eingeblendete schwarze Zwillingspunkt unterhalb, dass keine Zweifel mehr gerechtfertigt sind. Das gilt auch (gerade) noch für $p = 0{,}5$." Überprüfe NIKOs Aussage durch Nachrechnen und durch Kontrollen auch für einige andere Wahrscheinlichkeiten, dass das Programm korrekt arbeitet.

c) Durch Verschieben von p entsteht ein Intervall, das genau *diejenigen Wahrscheinlichkeiten enthält, die man nach der Beobachtung von h nicht zu bezweifeln braucht.*

d) Es heißt Konfidenzintervall zu h. Begründen Sie: Die Grenzen des Konfidenzintervalls sind die Lösungen der Gleichungen $0{,}4 = p \pm 2\frac{\sqrt{p\cdot(1-p)}}{\sqrt{n}}$.

3.2.3 Dualität – Simulation

HANNAH: „Eigentlich schließt man mit Konfidenzintervallen bei einer Datenerhebung vom Versuchsumfang n und der relativen Häufigkeit h auf die unbekannte Wahrscheinlichkeit p.

In **Abb. 21** habe ich den Gedanken umgekehrt und

- die Wahrscheinlichkeit p (auf 0,35) festgelegt,
- hiermit viele relative Häufigkeiten h zu einem festen Stichprobenumfang n erzeugt,
- für jedes h das zugehörige Konfidenzintervall berechnen und zeichnen lassen,
- die Konfidenzintervalle eingefärbt, die p nicht überdecken."

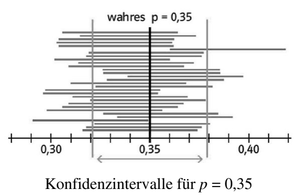


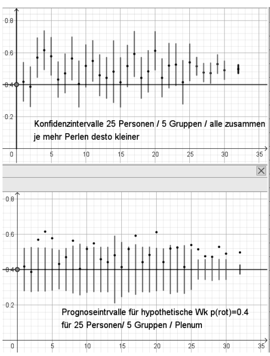
Abb. 21 *konfidenzintervalle-simultion.ggb*

a) Schätze nach Augenmaß die Größe n des in **Abb. 21** verwendeten Versuchsumfangs.
b) Untersucht mit dem Programm arbeitsteilig den Anteil der eingefärbten (also p enthaltenden) Konfidenzintervalle für $p = 0{,}2$ bzw. $p = 0{,}3$ bzw. $p = 0{,}5$ und $n = 25$ bzw. $n = 100$ bzw. $n = 400$. Formuliert eine Hypothese.
c) MAX: „Ist doch klar, was da herauskommen muss:
   (i) Ich weiß: Die relative Häufigkeit h liegt mit Wahrscheinlichkeit $95\ \%$ im Prognoseintervall zu p (vgl. Doppelpfeil in **Abb. 21**),
   (ii) deswegen ist das Konfidenzintervall mit 95 % Wahrscheinlichkeit gefärbt,
   (iii) das gilt für jedes n und für jedes p.“
   Vergleiche die Aussagen von MAX mit deinen Simulationsergebnissen.
d) Begründe (ii) und (iii), dass MAX recht hat. Nutze dazu die Definition (die „Dualitätsaussage“): „p liegt genau dann im Konfidenzintervall von h, wenn h im Prognoseintervall von p liegt.“

3.2.4 Dualität – Realexperiment

Abb. 22 zeigt eindrucksvoll, wie man die Dualität zwischen Prognose- und Konfidenzintervallen auch in einem Realexperiment mit tatsächlich „händisch gestochenen“ Stichproben erleben kann: 25 Löffelstichproben aus einem Topf mit roten, gelben und grünen Perlen in unbekannter Zusammensetzung wurden in den Spalten C – F ausgezählt. Die obere Grafik zeigt die 25 Konfidenzintervalle für den Anteil p der roten Perlen der 25 Versuchsteilnehmer. Sie sind umso kürzer, je mehr Perlen mit dem Löffel gestochen wurden (Spalte F). Beim Zusammenfassen in Fünfergruppen (F8, F15, F22 …) oder im Plenum (mit 1398 Perlen) verkleinern sie sich drastisch. 8 der 25 Konfidenzintervalle überdecken $p = 0{,}4$ nicht (Zellen J1 und G1) – und keines der Gruppenkonfidenzintervalle überdeckt $p = 0{,}4$. Der untere Teil der Grafik zeigt – dual dazu – die Prognoseintervalle, die zur Hypothese $p = 0{,}4$ und den 25 unterschiedlichen Stichprobengrößen gehören. Auch hier liegen genau 8 der 25 „gestochenen“ relativen Häufigkeiten außerhalb der Prognoseintervalle, ebenso alle 5 „Gruppenhäufigkeiten“. Die Entschlüsselung der Zusammenhänge zwischen den beiden Grafiken lässt durch Variierbarkeit der Hypothesenwahrscheinlichkeit p die Einsicht in die Dualität zwischen Konfidenz und Prognose „wie Schuppen von den Augen“ fallen. Jedem Lernenden und jeder Gruppe „gehört“ sowohl ein Konfidenz- wie auch ein Prognoseintervall. Dadurch erzeugt dieses Realexperiment neben Einsicht in Zusammenhänge ein persönliches „Eingebundensein“, wie sie keine noch so pfiffige Simulation bietet.

Abb. 22 Echte Stichproben



	A	..	C	D	E	F	G	H	I	J	K	L
1		..	rote	gel…	grü…	Su…	0.4	<-p…	5	8		
2							von	bis	h	zweifel	k-von	k-bis
3	1	..	28	22	17	67	0.28	0.52	0.42	0	0.3	0.54
4	2	..	24	16	22	62	0.28	0.53	0.39	0	0.26	0.51
5	3	..	37	10	18	65	0.28	0.52	0.57	1	0.45	0.69
6	4	..	40	15	10	65	0.28	0.52	0.62	1	0.49	0.74
7	5	..	41	15	15	71	0.28	0.52	0.58	1	0.46	0.69
8	26	..	170	78	82	330	0.35	0.45	0.52	1	0.46	0.57
9												
10	6	..	35	24	22	81	0.29	0.51	0.43	0	0.32	0.54
11	7	..	32	17	19	68	0.28	0.52	0.47	0	0.35	0.59
12	8	..	31	8	16	55	0.27	0.53	0.56	1	0.43	0.7
13	9	..	19	13	15	47	0.26	0.54	0.4	0	0.26	0.55
14	10	..	30	18	10	58	0.27	0.53	0.52	0	0.39	0.65
15	27	..	147	80	82	309	0.34	0.46	0.48	1	0.42	0.53
16												
17	11	..	34	17	11	62	0.28	0.53	0.55	1	0.42	0.67
18	12	..	28	11	22	61	0.28	0.53	0.46	0	0.33	0.59
19	13	..	27	19	15	61	0.28	0.53	0.44	0	0.32	0.57
20	14	..	13	6	8	27	0.21	0.59	0.48	0	0.29	0.67
21	15	..	17	9	15	41	0.25	0.55	0.41	0	0.26	0.57
22	28	..	119	62	71	252	0.34	0.46	0.47	1	0.41	0.54
23												

Abb. 23 Realexperiment: Dualität beim Perlenstechen

3.2.5 Strategievergleich

Es soll geprüft werden, ob $p = 0{,}8$ im Konfidenzintervall zur relativen Häufigkeit $h = 15\,/\,25 = 0{,}6$ liegt ($n = 25$).

1. Löse $x \pm 2\sqrt{x\cdot(1-x)/25} = 0{,}6.$ 2. Berechne $0{,}8 \pm 2\sqrt{0{,}8\cdot(1-0{,}8)/25}.$

a) Erläutere, welche Gedanken hinter den Lösungsansätzen (1) und (2) stecken und führe die Prüfung auf beiden Wegen durch.
b) Begründe, welchen Lösungsweg du wählen würdest, wenn nur ein einfachster Taschenrechner zur Verfügung steht.

4 Resümee

Wenn man im Sinne KOLMOGOROFFs Wahrscheinlichkeiten in all ihrem Facetten als Modelle („Hypothesen“) deutet, welche die Wirklichkeit besser oder schlechter beschreiben, nie aber ganz genau, ergibt das Schwarz-Weiß-Denken des Annehmens und Verwerfens beim Signifikanztest wenig Sinn. Es kann immer nur um besser oder schlechter gehen, um weniger oder mehr Zweifel. Und das $1/\sqrt{n}$ -Gesetz gestattet einen behutsamen Rückschluss auf die Stärke des Zweifels. Die Bedeutung, die dabei dem Versuchsumfang zukommt, wird aber erst beim Blick auf die Länge des zugehörigen Konfidenzintervalls sichtbar: Wenn bei sehr großem Versuchsumfang eine angenommene Wahrscheinlichkeit außerhalb des dann sehr kleinen Konfidenzintervalls liegt, ist das möglicherweise so, als würde man im Baumarkt (**Abb. 24**) die Länge eines zugesägten Brettes mit dem Mikroskop – also auf einer völlig unangemessenen Präzisionsskala – nachmessen. In der Regel wird man das Brett an der Kasse nicht reklamieren und man kann auch mit der angenommenen Wahrscheinlichkeit hervorragend weiterarbeiten. Das statistische Werkzeug der Konfidenzintervalle kommt damit dem gesunden Menschenverstand sehr viel weiter entgegen als der Hypothesentest. Möge die Vernunft auch in den Kreisen der Lehrplankommissionen siegen.

Abb. 24

Literatur

[1] AMRHEIN, V. GREENLAND, S. MCSHANE, B. [2019]: Retire statistical significance. Nature 567, S. 305–307. 2019.
[2] BRANDT, D. HOCHE, D., RIEMER, W. WOLLMANN, W. [2018]: Lambacher-Schweizer, Themenband Stochastik – Hessen. Stuttgart 2018. Klett Verlag.
[3] FREUDENTHAL, H. [1972]: The empirical law of large numbers or the stability of frequencies. In: Educational Studies in Mathematics, 4, S. 484–490.
[4] HUMENBERGER, H. [2019]: Der „empirische Wahrscheinlichkeitsbegriff“ – gut gemeint, aber auch wirklich gut? Stochastik in der Schule 39 (3), 20–24.
[5] LERGENMÜLLER, A., SCHMIDT, G., KRÜGER, K. (Hrsg.) [2012]: Neue Wege Stochastik. Braunschweig 2012.
[6] HENZE, N. [2017]: Stochastik für Einsteiger. Berlin, Springer.
[7] RIEMER, W. [1991]: Das $1/\sqrt{n}$-Gesetz. Stochastik in der Schule, 11(3), 24–36.
[8] RIEMER, W. [2019]: Konfidenzintervalle. Handreichung im Thüringer Schulportal. https://www.schulportal-thueringen.de/web/guest/media/detail?tspi=3498.
[9] RIEMER, W. [2019]: Grundvorstellungen beurteilender Statistik. MU 6/2019 S. 11.21.
[10] SACHS, L. [1999]: Angewandte Statistik. Berlin 1999, 9. Aufl. Springer Verlag.
[11] VEHLING, R. [2011]: Mit Simulationen zum Konfidenzintervall: PM 52 (39) S. 25–31.

Die hier erwähnten Programme können bei den Autoren per mail angefordert werden.

REIMUND VEHLING

Ein kleiner Blick auf Konfidenzintervalle in niedersächsischen Abituraufgaben

1 Einleitung

In Niedersachsen gehört das Testen von Hypothesen im Zentralabitur seit 2011 der Vergangenheit an. Stattdessen werden Aufgaben zu Konfidenzintervallen gestellt. Lehrkräfte und Lernende haben diesen Umstieg überwiegend positiv aufgenommen. Entscheidend für diesen Erfolg waren die umfangreichen und frühzeitig durchgeführten Fortbildungen. Hierbei ging es im Wesentlichen um inhaltliche Fragestellungen wie zentrale Einstiege, Rechnereinsatz, Aufgaben zum Entdecken, Üben, Festigen und Vertiefen (vgl. RIEMER/VEHLING [2020] in diesem Heft) sowie um die Planung konkreter Unterrichtseinheiten.

Im Folgenden geht es nicht darum, die gestellten Abituraufgaben einer kritischen Analyse zu unterziehen. Wir beschränken uns auf einen kommentierenden Überblick über die bisher verwendeten Aufgabentypen.

Die Aufgaben werden nach einer gewissen Zeit den Lehrkräften zur Verfügung gestellt, aber nicht veröffentlicht. Deshalb werden in diesem Artikel auch keine Originalaufgaben verwendet. Wegen der austauschbaren Kontexte muss das aber kein Nachteil sein. Natürlich sind die viel geschmähten „BlaBla“-Aufgaben kein Alleinstellungsmerkmal von Signifikanztests, es gibt sie (nicht nur) im Abitur selbstverständlich auch bei Konfidenzintervallen: Tatsächlich müssen bei Berechnungsaufgaben oft nur die Stellen gesucht werden, an denen die Informationen zum Stichprobenumfang, zum Stichprobenergebnis sowie zur Sicherheitswahrscheinlichkeit stehen. Werden diese drei Werte dann in den Rechner eingegeben, ist man fertig. Aufgaben dieses Formats sind – bei einer gewissen kritischen Distanz – als Routineaufgaben zur Überprüfung von Fertigkeiten durchaus sinnvoll, aber es darf nicht dabei bleiben. Es muss auch Aufgaben geben, die das Verständnis von Zusammenhängen überprüfen. Wie im Folgenden dargelegt, geschah dies in Niedersachen bisher u. a. durch folgende Aufgabentypen:

1. Berechnen eines Konfidenzintervalls
   a) ohne Interpretation
   b) mit Interpretation (Entscheidung treffen; Vermutung beurteilen)
2. zwei Stichprobenergebnisse zu einem Konfidenzintervall zusammenführen
3. Umkehraufgaben: linke oder rechte Grenze ermitteln
4. Zuordnungen mit Begründungen vornehmen
   a) mehrere Konfidenzellipsen
   b) verschiedene, je n-mal dargestellte Konfidenzintervalle
   c) verschiedene, je n-mal dargestellte Häufigkeitsverteilungen der linken oder rechten Grenzen von Konfidenzintervallen
5. Unterschiede zwischen Prognose- und Konfidenzintervallen reflektieren und erläutern
6. Vernetzung von Binomialverteilungen mit Konfidenzintervallen

Aufgaben zur Ermittlung eines erforderlichen Stichprobenumfangs wurden bisher nicht gestellt.

Da unterschiedliche Berechnungsmethoden und Darstellungen existieren (vgl. HENZE [2020] in diesem Heft), werfen wir vorab einen Blick auf zentrale unterrichtsrelevante Aspekte. Dabei beschränken wir uns auf die Sicherheitswahrscheinlichkeit $\gamma = 95\%$, also auf die $1{,}96\sigma$-Intervalle. Und wir gehen ein auf den Unterschied zwischen dem zufälligen Intervall als Bereichsschätz*verfahren* und deren Realisationen, also konkreten zufallsabhängigen Intervallen. Tatsächlich wird in allen Schulbüchern und auch in den Aufgaben zum Zentralabitur eine Realisation mit dem Begriff *Konfidenzintervall* gleichgesetzt. In meinem Unterricht wird der Unterschied thematisiert.

2 Ein Blick in meinen Unterricht

Die Berechnung eines Konfidenzintervalls ist einfach, viel schwieriger ist es in der Schule, einen Lernweg so zu gestalten, dass ein Verständnis vom Prinzip des Schätzens aufgebaut wird. Leider ist die Interpretation des per Knopfdruck erstellten Ergebnisses alles andere als einfach. Die Lehrperson sollte den zentralen Unterschied zwischen einem Konfidenzintervall als Bereichs-Schätzverfahren und den konkreten Konfidenzintervallen (Realisierungen) kennen (vgl. HENZE [2020] in diesem Heft). In Niedersachsen hat sich ein Zweischritt etabliert: Zuerst die Berechnung der Intervalle und dann mithilfe von Simulationen die Interpretation der Sicherheitswahrscheinlichkeit. Wenn hier keine klare Trennung vorgenommen wird, können Lernende nicht verstehen, warum die Aussage „*p wird mit 95 %iger Wahrscheinlichkeit in dem berechneten Intervall liegen*" falsch ist: Nur die Grenzen des zufälligen Intervalls sind Zufallsgrößen, nicht p. Entweder ist der unbekannte Anteil p im Intervall enthalten oder nicht. Von einer Wahrscheinlichkeit zu sprechen, mit der p im Intervall liegt, ergibt somit keinen Sinn.

In der Schule wird die Zufälligkeit des konkreten Konfidenzintervalls durch Simulationen erlebbar: Zu einem (in der Realität unbekannten, in der Simulation aber festlegbaren) p werden z. B. 100 Konfidenzintervalle durch eine Simulation erzeugt. Dann wird die Anzahl der Intervalle gezählt, die p überdecken. So führt z. B. die Sprechweise „*ca. 95 % aller so konstruierten Intervalle überdecken p*" zu einer sinnvollen Interpretation.

Eine weitere wichtige Reduktion stellt der Begriff des Prognoseintervalls dar. Eigentlich handelt es sich um eine hochwahrscheinliche Menge: Hier ist die Approximation der Binomialverteilung durch die Normalverteilung enthalten. Weiterhin bieten Prognoseintervalle die Möglichkeit, handlungsorientiert und anschaulich die Berechnung eines Konfidenzintervalls einzuführen, etwa so: Alle Werte für p, in deren 95 %-Prognoseintervall das Stichprobenergebnis h liegt, bilden das 95 %-Konfidenzintervall. Es handelt sich hierbei um das WILSON-Konfidenzintervall. Um die Grenzen zu erhalten, müssen die beiden Gleichungen

$$h = p - 1{,}96 \cdot \sqrt{\frac{p \cdot (1-p)}{n}} \quad \text{und} \quad h = p + 1{,}96 \cdot \sqrt{\frac{p \cdot (1-p)}{n}} \tag{1}$$

nach p aufgelöst werden. Diese Berechnungsmethode wurde in den ersten Jahren im Zentralabitur häufig angewendet. Aufgaben im Umfeld der Konfidenzellipse wurden erst später gestellt. Im Kerncurriculum wird diese grafische Veranschaulichung nicht erwähnt.

In allen neueren Schulbüchern finden sich an zentraler Stelle Darstellungen wie in **Abb. 1** und **Abb. 3**. Die Veranschaulichungen sollten in keinem Unterricht fehlen. **Abb. 1** kann handlungsorientiert entwickelt werden und zeigt beeindruckend das zugrunde liegende Berech-

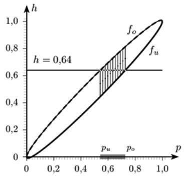


Abb. 1 Konfidenzellipse mit dem Konfidenzintervall

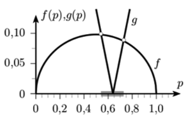


Abb. 2 Ellipse, Betragsfunktion und Konfidenzintervall

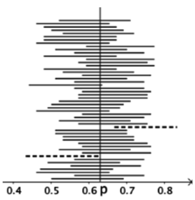


Abb. 3 Simulationen von Konfidenzintervallen

nungsverfahren eines Konfidenzintervalls auf. Mehr noch: Diese Darstellung zeigt deutlich den Unterschied zwischen Prognoseintervallen (vertikale Strecken) und einem Konfidenzintervall (horizontale Strecke). **Abb. 3** führt direkt auf die Interpretation der Sicherheitswahrscheinlichkeit. **Abb. 2** wird selten benutzt. Hierzu gehört der Ansatz $|h-p| = 1{,}96 \cdot \sqrt{p(1-p)/n}$. Mithilfe der Grafik kann gut erkannt werden, wieso das zugehörige Konfidenzintervall (WILSON) nur für $h = 0{,}5$ symmetrisch zu h ist.

Mit der Substitution $a := \frac{1}{1+\frac{1{,}96^2}{n}}$ erhält man als Lösung der beiden Gleichungen (1) die beiden Grenzen p_u und p_o:

$$p_{u,o} = (1-a)\cdot\frac{1}{2} + a\cdot h \mp a\cdot 1{,}96\cdot\sqrt{\frac{1}{4n}\cdot\frac{1-a}{a}+\frac{h\cdot(1-h)}{n}}. \qquad (2)$$

Interessant: Der Mittelwert des WILSON-Konfidenzintervalls ist das gewichtete Mittel aus dem Stichprobenergebnis h und 0,5. Diese Darstellung zeigt übrigens auch sofort, dass nur für $h = 0{,}5$ eine Symmetrie zu h vorliegt.

Schon für $n = 400$ ist $a = 0{,}990\ldots \approx 1$. Damit erhält man ein zweites approximatives Intervall, das sogenannte WALD-Konfidenzintervall:

$$\left[h - 1{,}96\cdot\sqrt{h(1-h)/n};\ h + 1{,}96\cdot\sqrt{h(1-h)/n}\right]. \qquad (3)$$

In vielen Rechnern wird die Berechnung des WALD-Konfidenzintervalls auf Knopfdruck zur Verfügung gestellt (Bsp.: 1−PropZInt). Die Berechnung eines Konfidenzintervalls kann im Zentralabitur auch mit dieser Näherung durchgeführt werden, auch wenn die Überdeckungswahrscheinlichkeit für p-Werte nahe 0 oder 1 deutlich kleiner als die Vorgabe ist. Nach der Einführung eines Konfidenzintervalls nach WILSON sollte m. E. Formel (3) trotz der Probleme am Rand auch eingesetzt werden. Auf die Verwechslungsgefahr mit einem Prognoseintervall muss eingegangen werden. Beide Intervalle haben die gleiche Struktur.

3 Die Abituraufgabentypen

Wie angekündigt werfen wir nur einen kommentierenden Blick auf die bisher in Niedersachsen verwendeten Aufgabenformate:

3.1 Typ 1: Berechnen eines Konfidenzintervalls

3.1.1 … ohne Interpretation

Aufgabe 1: Im Rahmen einer Studie gaben 242 von 550 Jugendlichen eines Bundeslandes an, dass sie nicht frühstücken, bevor sie in die Schule gehen.
Bestimmen Sie auf Basis dieser Studie ein Konfidenzintervall zu einer Sicherheitswahrscheinlichkeit von 95 %.

Lösung: Es gilt $h = \frac{242}{550} = 0{,}44$.

Um die Grenzen des 95%-WILSON-Konfidenzintervalls zu bestimmen, müssen die Gleichungen $0{,}44 = p + 1{,}96 \cdot \sqrt{\frac{p \cdot (1-p)}{550}}$ und $0{,}44 = p - 1{,}96 \cdot \sqrt{\frac{p \cdot (1-p)}{550}}$ gelöst werden.

Die Lösungen werden mithilfe des Rechners durch Schnittpunktbestimmung ermittelt. Es ergibt sich das 95%-WILSON-Konfidenzintervall $[0{,}3990\ldots;0{,}4817\ldots]$. Das 95%-WALD-Konfidenzintervall $[0{,}3985\ldots;0{,}4814\ldots]$ unterscheidet sich hiervon nur unwesentlich.

Bemerkungen: Die Wahl des unbestimmten Artikels in der Abituraufgabe ist sinnvoll, da es mehrere Berechnungsmöglichkeiten für Konfidenzintervalle gibt.

Im Erwartungshorizont wurden bis 2019 keine gerundeten Werte angegeben, sondern eine Darstellung mit Punkten (s. Lösung) gewählt.

3.1.2 … mit Interpretation

Aufgabe 2: Kandidat A behauptet vor einer Bürgermeisterwahl, mindestens 50 % der Stimmen zu erhalten. Er lässt eine Umfrage unter 750 Wahlberechtigten durchführen. Hierbei geben 345 Personen an, den Kandidaten A wählen zu wollen.

Entscheiden Sie mithilfe eines Konfidenzintervalls, ob man bei einer Sicherheitswahrscheinlichkeit von $\gamma = 95\,\%$ davon ausgehen kann, dass Kandidat A mindestens 50 % der Stimmen erhält.

Lösung: Mit dem Stichprobenumfang $n = 750$ und der relativen Häufigkeit $h = 0{,}46$ erhält man das 95%-WILSON-Konfidenzintervall [0,4246…; 0,4957…]. Da die rechte Intervallgrenze kleiner als 0,5 ist, überdeckt dieses (konkrete) Konfidenzintervall nicht 0,5. Die Behauptung des Kandidaten, mindestens 50 % der Stimmen zu erhalten, ist mit dem Ergebnis dieser Umfrage nicht verträglich.

Für die rechte Intervallgrenze des 95%-WALD-Konfidenzintervalls ergibt sich der Wert 0,4956…, der auch kleiner als 0,5 ist.

3.2 Typ 2: Zwei Stichprobenergebnisse zu einem Konfidenzintervall zusammenführen

Aufgabe 3: Bei einer Umfrage wurden 1000 Personen befragt, welches Betriebssystem (A oder B) sie benutzen. Von diesen gaben 354 das Betriebssystem A und 264 das System B an, 382 gaben ein anderes Betriebssystem an.

Bestimmen Sie jeweils ein Konfidenzintervall ($\gamma = 95\,\%$) für den Anteil der Personen, die Betriebssystem A benutzen, Betriebssystem B benutzen, entweder A oder B benutzen.

Vergleichen Sie das zuletzt bestimmte Konfidenzintervall mit dem Intervall, das sich durch Addition der jeweiligen Intervallgrenzen aus den beiden zuvor bestimmten einzelnen Konfidenzintervallen ergibt.

Lösung: Beispielhaft werden hier WALD-Konfidenzintervalle benutzt. Für A ergibt sich damit das Konfidenzintervall $[0{,}3243\ldots;0{,}3836\ldots]$, für B ergibt sich $[0{,}2366\ldots;0{,}2913\ldots]$. Durch Addition der jeweiligen Intervallgrenzen resultiert das Intervall [0,5610…; 0,6749…]. Beide zusammen: 618 von 1000 Befragten (A oder B) liefert [0,5878…; 0,6481…]. Offensichtlich ergibt sich dieses nicht aus der Addition der Intervallgrenzen der Einzelintervalle.

Bemerkung: Es wurde nach keiner Begründung gefragt. Das ist auch für eine Abituraufgabe sehr sinnvoll. Im Unterricht kann man hieraus einen Forschungsauftrag machen: Führt die Addition der Intervallgrenzen immer zu einem „zu großen“ Konfidenzintervall? Das hat nichts mehr mit Stochastik zu tun, sondern „nur noch“ mit Algebra.

3.3 Typ 3: Umkehraufgabe: Grenze gesucht

Aufgabe 4: Eine Umfrage unter 1000 Wahlberechtigten liefert für Partei A das 95 %-Konfidenzintervall $[0,4026;b]$. Hierbei ist der Wert für die linke Intervallgrenze auf vier Nachkommastellen gerundet.

Bestimmen Sie den Wert von *b*.

Lösung (Endergebnis auf 4 Nachkommastellen gerundet):

Mithilfe des 95%-WALD-Konfidenzintervalls liefert $0,4026 = h - 1,96 \cdot \sqrt{\frac{h \cdot (1-h)}{1000}}$ die Lösung $h^* = 0,4333\ldots$ Der Wert für b ergibt sich zu $b = h^* + 1,96 \cdot \sqrt{\frac{h^* \cdot (1-h^*)}{1000}} = 0,4640.$

Ermittlung des Wertes für b mit WILSON-Konfidenzintervallen:

Für den Stichprobenanteil h gilt: $h^* = 0,4026 + 1,96 \cdot \sqrt{\frac{0,4026 \cdot (1-0,4026)}{1000}} = 0,43299\ldots$

Mit $h^* = \mathrm{b} - 1,96 \cdot \sqrt{\frac{\mathrm{b} \cdot (1-\mathrm{b})}{1000}}$ folgt $\mathrm{b} = 0,4639.$

Bemerkungen: Dieser Aufgabentyp wurde schon oft von mir im Unterricht eingesetzt. Vor der Einführung von WALD-Konfidenzintervallen fiel den Lernenden dieses Problem schwer. Das Zeigen der Abbildung einer Konfidenzellipse führte dann aber schnell zu einer Lösungsidee. Das Vorgehen zeigt zwei wichtige Punkte auf: Mit dieser Umkehraufgabe kann zum einen diagnostiziert werden, ob ein bekanntes Verfahren in einer neuen Situation angewendet werden kann. Zum anderen zeigt dieses Problem, wie hilfreich ein Darstellungswechsel zur Problemlösung sein kann. Eine Reflexion des Vorgehens ist (nicht nur) an dieser Stelle sinnvoll.

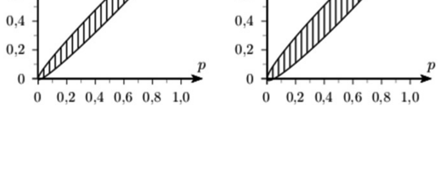
In einer Abituraufgabe nach Vorgabe der linken Grenze, der Stichprobenanzahl *n* sowie dem Stichprobenergebnis *k* wurde nach der rechten Intervallgrenze gefragt. Zur Lösung muss zuerst der Faktor *z* zur Sicherheitswahrscheinlichkeit ermittelt werden. Für die beiden Näherungsintervalle ergeben sich aber unterschiedliche Sicherheitswahrscheinlichkeiten. Es sollte deshalb in der Aufgabenstellung die zu benutzende Methode vorgegeben werden.

3.4 Typ 4: Zuordnungen mit Begründungen vornehmen

Aufgabe 5: Mit den beiden folgenden Abbildungen können Konfidenzintervalle zur Sicherheitswahrscheinlichkeit $\gamma = 0,95$ für $n = 75$ bzw. für $n = 150$ grafisch ermittelt werden.

Entscheiden Sie, welche der Darstellungen zu $n = 75$ bzw. zu $n = 150$ gehört.

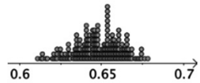
Lösung: In den Abbildungen werden für verschiedene Werte von p die zugehörigen 95%-Prognoseintervalle dargestellt. Hiermit können Konfidenzintervalle grafisch ermittelt werden, da alle Werte für *p*, in deren 95%-Prognoseintervall das Stichprobenergebnis h liegt, das 95%-Konfidenzintervall bilden.



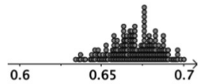
Diese Intervalle haben jeweils die Länge $2 \cdot 1,96 \cdot \sqrt{p(1-p)/n}$. Die Länge ist also proportional zu $1/\sqrt{n}$. Die Intervalle zu $n = 75$ sind damit um den Faktor $\sqrt{2}$ größer als die entsprechenden Intervalle zu $n = 150$. Somit gehört Abbildung *A* zu $n = 150$ und Abbildung *B* zu $n = 75$.

Aufgabe 6: Es werden 200 Stichproben mit $n = 100$ simuliert. Für diese werden jeweils die zugehörigen Konfidenzintervalle für die beiden Sicherheitswahrscheinlichkeiten 70 % und 99 % berechnet.

Die nebenstehenden Abbildungen zeigen als Punktdiagramm jeweils die rechten Intervallgrenzen der zugehörigen Konfidenzintervalle.



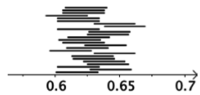
Entscheiden Sie, welche der beiden Abbildungen zur Sicherheitswahrscheinlichkeit 70 % und welche zur Sicherheitswahrscheinlichkeit 99 % gehört.



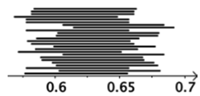
Lösung: Je kleiner die Sicherheitswahrscheinlichkeit ist, desto kleiner sind die zugehörigen Konfidenzintervalle. Deshalb gehört die obere Häufigkeitsverteilung der rechten Intervallgrenzen zu $\gamma = 70\,\%$, da sie weiter links in Richtung des unbekannten Parameters p liegt.

Aufgabe 7: Es werden 25 Stichproben mit $n = 100$ simuliert. Für diese werden jeweils die zugehörigen Konfidenzintervalle für die beiden Sicherheitswahrscheinlichkeiten 70 % und 99 % berechnet.

Die nebenstehenden Abbildungen zeigen jeweils 25 berechnete Konfidenzintervalle als Strecken übereinander.



Entscheiden Sie, welche der beiden Abbildungen zur Sicherheitswahrscheinlichkeit 70 % und welche zur Sicherheitswahrscheinlichkeit 99 % gehört.



Lösung: Je kleiner die Sicherheitswahrscheinlichkeit ist, desto kleiner sind die zugehörigen Konfidenzintervalle. Deshalb gehört die obere Abbildung zu $\gamma = 70\,\%$, da $0{,}7 < 0{,}99$ gilt.

3.5 Typ 5: Unterschiede zwischen einem Prognose- und einem Konfidenzintervall erläutern

Aufgabe 8: Erläutern Sie die unterschiedlichen Bedeutungen eines 95%-Prognoseintervalls und eines Konfidenzintervalls zur Sicherheitswahrscheinlichkeit 95 %.

Lösung: Bei der Bestimmung eines Prognoseintervalls ist die Wahrscheinlichkeit p bekannt. Die relative Häufigkeit ist vom Zufall abhängig und liegt mit einer Wahrscheinlichkeit von etwa 95 % in dem Prognoseintervall. Wahrscheinlichkeiten werden als Prognose für relative Häufigkeiten interpretiert. *Man schließt von der Grundgesamtheit auf die Stichprobe, von der Wahrscheinlichkeit auf die relative Häufigkeit, vom Modell auf die Realität.*

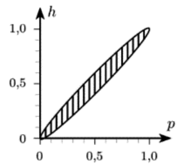
Ein Konfidenzintervall zur Sicherheitswahrscheinlichkeit 95 % dagegen wird mithilfe der relativen Häufigkeit einer Stichprobe bestimmt. Hierbei ist p fest, aber unbekannt. Die Grenzen des Konfidenzintervalls hängen vom Zufall ab, genauer: Die Grenzen sind Realisationen des zugehörigen Schätzverfahrens. Das (konkrete) Konfidenzintervall enthält alle Wahrscheinlichkeiten, in deren 95%-Prognoseintervall das Stichprobenergebnis liegt. *Man schließt von der Stichprobe auf die Grundgesamtheit, von der relativen Häufigkeit auf die Wahrscheinlichkeit, von der Realität auf das Modell.*

3.6 Typ 6: Vernetzung von Binomialverteilung und Konfidenzintervall

Aufgabe 9: Es werden 25 Stichproben vom Umfang $n = 20$ mit einer vorgegebenen (eigentlich unbekannten) Wahrscheinlichkeit p simuliert. Für jede Stichprobe soll ein Konfidenzintervall zur Sicherheitswahrscheinlichkeit $\gamma = 90\,\%$ für die (unbekannte) Wahrscheinlichkeit p berechnet werden.

Bestimmen Sie die Wahrscheinlichkeit dafür, dass alle 25 Konfidenzintervalle die (unbekannte) Wahrscheinlichkeit p überdecken.
Lösung: Die Zufallsgröße X, die die Anzahl der Konfidenzintervalle angibt, die p überdecken, kann als binomialverteilte Zufallsgröße modelliert werden mit $n = 25$ und $p = 0{,}90$. Daraus folgt für die gesuchte Wahrscheinlichkeit: $P(X = 25) = 0{,}9^{25} \approx 0{,}07$.

4 Beispiel einer mündlichen Prüfung zum Thema Konfidenzintervalle



Das Schätzen kann natürlich auch in einer mündlichen Abiturprüfung zur Sprache kommen. Das folgende Problem habe ich in einer mündlichen Abiturprüfung benutzt:

1. Jemand wirft 100-mal eine Münze und erhält 42-mal Kopf.
Äußern Sie sich zu der Behauptung: *Die Münze muss gezinkt sein.*
2. Es wird auf einer Folie eine Grafik (s. links) präsentiert ($n = 100$; $\alpha = 0{,}95$). Daran soll aufgezeigt werden, wie man damit für die Problemstellung in (1) zu einem Konfidenzintervall gelangt.
3. Mögliche Vertiefungen:
Welche Auswirkungen ergeben sich, wenn die Stichprobenanzahl n verkleinert/vergrößert wird (γ verkleinert/vergrößert wird)?
Eingehen auf das $1/\sqrt{n}$ -Gesetz
Unterschiede zwischen einem Prognose- und einem Konfidenzintervall

5 Fazit

Wie sich in Niedersachsen gezeigt hat, lassen sich bei Abiturprüfungen zu Konfidenzintervallen vielfältige Problemstellungen auch jenseits der „*BlaBla-Aufgaben*“ finden. Ich würde es begrüßen, wenn die beiden Methoden (WILSON, WALD) verbindlich vorgegeben werden. WALD-Konfidenzintervalle können (wie in Österreich) als symmetrische Konfidenzintervalle bezeichnet werden. Dann könnten die Aufgaben noch klarer gestellt werden. Weiterhin ergeben sich gerade bei der Betrachtung symmetrischer Konfidenzintervalle durch die einfachere Struktur mehr Möglichkeiten, Verständnisfragen zu stellen. Die Tatsache, dass diese Intervalle schnell mit Prognoseintervallen verwechselt werden können, hat etwas Gutes: Ein Eingehen auf die Unterschiede zwischen diesen beiden Intervallen wird im Unterricht zwingend erforderlich. WALD-Konfidenzintervalle führen ohne Rechnereinsatz schnell zu einer ersten Abschätzung über die Größe des Konfidenzintervalls.

Ein besonders wichtiger Aspekt sei noch ergänzt: In der Realität gibt es keine repräsentativen „Zufalls“-Stichproben. Dieses Problem haben alle Meinungsforschungsinstitute. Nach meinen Informationen werden WALD-Intervalle benutzt. Vor dem Faktor 1,96 wird aber noch ein „*Design-Faktor*“ größer als 1 eingebaut. Damit werden die Intervalle größer, um die Überdeckungswahrscheinlichkeit möglichst einzuhalten. Darauf sollte im Unterricht auch eingegangen werden. Die Berechnungen kann gerne ein Rechner übernehmen.

Literatur

[1] HENZE, N. [2020]: Konfidenzbereiche für das p der Binomialverteilung – MU 4/2020.
[2] RIEMER, W., VEHLING, R. [2020]: Prognose- und Konfidenzintervalle: beurteilende Statistik mit Sinn und Verstand. MU 4/2020.