Wo liegen die Knackpunkte stochastischer Begriffe und Modelle?

MANFRED BOROVCNIK, KLAGENFURT

Der Aufsatz thematisiert die Herausforderungen und Schwierigkeiten, die beim Verständnis und der Anwendung stochastischer Konzepte und Modelle auftreten. Beleuchtet wird insbesondere die Diskrepanz zwischen der vermittelten Theorie und der tatsächlichen Anwendung sowie das häufige Missverständnis und die Ablehnung, denen die Stochastik sowohl in der Bildung als auch in der öffentlichen Wahrnehmung begegnet.

Ein zentraler Punkt ist dabei die kritische Reflexion über den Unterricht in Stochastik, der trotz jahrzehntelanger Bemühungen oft ins Leere läuft. Viele Lernende haben, trotz intensiven Lernens, Schwierigkeiten, die Konzepte zu verinnerlichen und anzuwenden. Worin bestehen die Gründe für diese Ablehnung und mangelnde Akzeptanz? Welche Möglichkeiten gibt es, wie die Lehre der Stochastik verbessert und effektiver gestaltet werden kann? Ein weiterer wichtiger Aspekt ist die Diskussion über die öffentliche Wahrnehmung stochastischer Aussagen und Modelle. Der Aufsatz zeigt auf, wie häufig in der Praxis stochastische Methoden missverstanden oder falsch angewendet werden, etwa bei der Vorhersage wirtschaftlicher Entwicklungen, der Wahlforschung oder der Gesundheitsvorsorge. Argumentiert wird, dass ein tiefgreifenderes Verständnis und eine kritischere Auseinandersetzung mit stochastischen Methoden notwendig sind, um deren Potenzial voll ausschöpfen zu können.

Der Text schließt mit der Feststellung, dass die Erwartungen an stochastische Methoden oft zu hoch sind und dass diese Methoden nur so gut sein können wie die zugrunde liegenden Modelle. Er plädiert dafür, den Zugang zur Stochastik als eine von mehreren Methoden zu verstehen und die Knackpunkte zu identifizieren, die diesen Zugang erleichtern. Der Text ist eine Aufforderung zur Reflexion und Diskussion über die Rolle der Stochastik in Bildung und Gesellschaft.

1. Jahrzehnte der Didaktik der Stochastik – Eine Bilanz

Trotz jahrelangen Unterrichts scheint die Lehre der Stochastik ins Leere zu laufen, wenn sie überhaupt stattfindet. Die Zentralmatura hat die Situation zwar etwas verbessert, doch der öffentliche Umgang mit stochastischen Begriffen und Fragestellungen sowie mit Daten ist weiterhin beklagenswert. Warum gibt es so viel Ablehnung und Unwissenheit? Was macht die Stochastik so schwer verständlich? Und wie können Menschen von den Chancen dieser Methoden überzeugt werden? Wo genau liegen, mit einem Wort, die Knackpunkte für einen erfolgreichen Unterricht in Stochastik? Gibt es im Vergleich zur restlichen Mathematik Unterschiede in Wissen, Fertigkeiten, Einstellungen oder Ablehnung? Der Text soll zur kritischen Diskussion anregen.

1.1 Pessimistische Gedanken im Jahr 2024

Können stochastische Methoden wirklich besser genutzt werden? Gibt es einen Unterschied im Erfolg zwischen Einzelpersonen und größeren Stakeholdern wie Unternehmen, Interessensgruppen oder dem Staat? Dazu drei Beispiele: Beim Brexit-Referendum sagten die meisten Institute ein knappes Pro-EU-Ergebnis voraus, am Ende stimmten jedoch 51,9 % für den Brexit (Gelman, 2016). Die Vorhersagen für die Verbreitung von COVID-19 sowie die Wirksamkeit der Maßnahmen dagegen lagen oft daneben (Adam, 2020; Joffe, 2021; Mader & Rüttenauer, 2022; Agrawal et al., 2022). Bei den US-Wahlen 2016 wurden Trump als Außenseiter in den Umfragen gegen H. Clinton kaum Chancen eingeräumt (Gelman & Azari, 2017).

1.2 Spezifische Schwierigkeiten der Stochastik

Wie soll ein Schüler die Thematik verstehen, wenn selbst die klügsten Köpfe die Regelmäßigkeiten der Stochastik missverstehen? Cardano (1560) glaubte, dass bei sechs Würfen mit einem Würfel alle sechs Zahlen erscheinen müssen – dass dies nicht so ist, scheint Schicksal zu sein. Ein Problem ist, dass die

Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG), Heft 56, 2024, S. 1 - 19.

Stochastik über einen einzelnen Fall wenig aussagen kann. „Mein Schwiegervater musste sich einer Routineoperation unterziehen, die in den letzten 18 Jahren nie misslungen war. Doch bei ihm ging es schief.“ Kinder können mit solchen Unsicherheiten wenig anfangen – sie müssen denken, das sei alles nur erfunden.

2. Die Rolle von Daten in einer evidenzbasierten Gesellschaft

Erkenntnisse werden aus Daten ‚geschöpft‘. ‚Signifikante‘ Ergebnisse werden oft als Beweise angesehen, anstatt sie als Anlass für weiterführende substanzwissenschaftliche Forschungen zu nutzen.

Daten werden oft missverstanden und als absolute Fakten interpretiert. In Wahrheit sind Daten jedoch nur im Vergleich interpretierbar und nur aussagekräftig, wenn sie unter bestimmten Bedingungen erhoben wurden (Borovcnik, 2014). Flops in Projekten der angewandten Statistik sind meist auf Versäumnisse in der Systemanalyse vor der Datenproduktion zurückzuführen, wie beispielsweise die Definition von Merkmalen, die Abgrenzung der Zielgesamtheit oder das Studium von Confoundern (Störgrößen; Borovcnik & Terán, 2023).

Ein bekanntes Sprichwort lautet: „Traue keiner Statistik, die du nicht selbst gefälscht hast!“

2.1 Umgang mit Daten

In Lehrbüchern wird das Thema Manipulation häufig behandelt, allerdings oft, ohne die zugrunde liegende Struktur und die Techniken der Manipulation ausreichend zu erläutern. Wenig überraschend ist daher, dass auch heute noch in Zeitreihen ‚beliebige‘ Zeitpunkte als Referenz herangezogen werden, anstatt aktuelle Daten unter Berücksichtigung der langfristigen Variabilität und Saisonalität zu bewerten. Diese Vorgehensweise führt zu Aktionismus und der Entstehung vermeintlicher Kausalketten: Maßnahmen werden als erfolgreich bewertet, und es wird gefordert, weitere Maßnahmen zu ergreifen.

Die Definition von Merkmalen wird zudem häufig geändert, ohne dass dies die Interpretation der Zahlen beeinflusst: So werden etwa Arbeitslosenzahlen oder CT-Werte bei Covid-Tests interpretiert, während wichtige andere Daten oft gar nicht verfügbar sind. Beispielsweise werden in Diskussionen über Pensionsreformen Entscheidungen getroffen, ohne dass eine ausreichende Datenbasis vorhanden ist.

2.2 Evidence-based Society – Ein Beispiel aus Österreich

Oft wird die günstigste Vergleichsbasis gewählt:

In Österreich gab es 2020 ein ‚Budgetloch‘ von 30 Milliarden Euro, das bis 2023 auf 12 Milliarden Euro geschrumpft ist – was 2,7 % des BIP entspricht. Gleichzeitig sind jedoch die Staatsschulden von 280 Milliarden Euro (2019) auf 371 Milliarden Euro in 2023 angestiegen (Statistik Austria, o.D.).

Wenn Daten als Grundlage für Entscheidungen dienen sollen, bleiben Definitionsprobleme oft ungelöst. So gibt es keine eindeutige Bilanz über die Staatsschulden, weil man kreativ vermeidet, Schulden überhaupt als Schulden zu zählen. Dies zeigt die Herausforderungen einer Evidence-based Society.

3. Grundsätzliche Missverständnisse

3.1 Daten sind Fakten

Von wegen (siehe Borovcnik, 2014). Das bedeutet jedoch keineswegs, dass man sie ‚verhandeln‘ kann. Der Mythos der Repräsentativität von Daten ist eine zentrale Grundlage der Stochastik. Tatsächliche

Relevanz von Daten zu erreichen und zu überprüfen ist jedoch äußerst schwierig. Es ist wenig hilfreich zu behaupten, dass es sich dann eben nicht um ‚echte' Daten handelt. Daten existieren und werden – unabhängig von ihrer Qualität – als Grundlage für Argumentationen genutzt.

3.2 Rivalisierende Konzepte und Vorstellungen sind irrational

Die Magie des Zufalls fasziniert mehr als abstrakte Begriffe und wird oft zu Unrecht abgewertet. Das beginnt mit der Abwertung von Laplacens Dämon, der alles weiß und die Wahrscheinlichkeit dem Unwissenden überlässt (Laplace, 1812), und endet nicht mit Einsteins skeptischer Haltung zum Zufall in der Physik („Gott würfelt nicht", Heisenberg, 1969). Menschen spielen in Glücksspielen und wetten aus verschiedensten Gründen, oft erfolgreich – manchmal sogar gegen den Rat von Stochastikern. Das Herausfordern des Schicksals entspricht archetypischen Entwicklungsmustern und ist eine wichtige Triebfeder für Fortschritt (Borovcnik, 2016a)). Deshalb ist es wichtig, die Stochastik richtig zu positionieren: Was kann sie leisten und was nicht? Was leisten andere Ansätze?

3.3 Perspektive des Individuums – Perspektive des Systems

Die Stochastik bietet für Systeme oft gute Lösungen, aber nicht unbedingt für das Individuum (Borovcnik, 2022). Die Frage „Ist das für mich relevant?" ist berechtigt. Im klassischen Versicherungsvertrag stellt sich die Frage: „Wie kann etwas gut für die Versicherung sein und gleichzeitig gut für mich?" Ein Arzt ist dem System verpflichtet und haftbar, während der Patient direkt die Folgen der Entscheidungen tragen muss. Diese Perspektivenfrage lässt sich auf jedes Glücksspiel, auf das Wettgeschäft und auch auf den Finanzmarkt übertragen. Die Black-Scholes-Gleichung ermöglicht die Bepreisung von Optionen und anderen Futures auf dem Finanzmarkt, liefert aber keine zuverlässigen Prognosen für den Ablauf einzelner Geschäfte (Black & Scholes, 1973).

3.4 Wahrscheinlichkeit als Eigenschaft von Objekten

Wahrscheinlichkeiten existieren und müssen lediglich gefunden und angemessen gemessen werden. Wahrscheinlichkeiten können an Objekte, Subjekte oder Szenarien gebunden sein. Szenarien unterscheiden sich von Modellen (Borovcnik, 2009): Modelle sollen möglichst gut passen, während Szenarien durchgespielt werden, um Möglichkeiten abzuschätzen. In vielen Anwendungen werden Eingangsgrößen nicht durch Experimente geschätzt, sondern als Szenarien angenommen. Bei subjektivistischen Wahrscheinlichkeiten ist die Revision von Urteilen in Form von bedingten Wahrscheinlichkeiten wichtig, bei objektivistischen Wahrscheinlichkeiten hingegen die Schätzung durch relative Häufigkeiten. Der Stellenwert bedingter Wahrscheinlichkeiten variiert entsprechend (Borovcnik, 2023).

3.5 Kleine Wahrscheinlichkeiten als ‚Business as Usual'

Kleine Wahrscheinlichkeiten werden oft so interpretiert, als wären zuverlässige Schätzungen darüber möglich, etwa durch Analysen von Netzwerken unter Annahme von Unabhängigkeit der Einzelteile. Aus Daten ließen sich solche Wahrscheinlichkeiten jedoch niemals verlässlich schätzen. Bis Mitte des 20. Jahrhunderts gab es Vorschläge, Wahrscheinlichkeiten unterhalb einer moralischen Grenze ganz wegzulassen. Nehmen wir an, es gehe um eine Wahrscheinlichkeit von 10^{-4}. Bei einer Zufallsstichprobe von 10.000 gibt es folgende Wahrscheinlichkeiten für die geschätzten Werte (siehe Tabelle 1). Die Werte zeigen, dass sehr große Schätzfehler eine relativ große Wahrscheinlichkeit haben, so ist ein Schätzfehler von 200 % oder mehr (Schätzwert größer gleich 0,0003) mit 8,1% Wahrscheinlichkeit ausgestattet und wird daher entsprechend häufig auftreten. Auf der anderen Seite hat der Schätzwert 0 fast 37 % Wahrscheinlichkeit und führt damit ganz in die Irre (siehe Borovcnik, 2022).

Tabelle 1: Verteilung der Schätzwerte für die ‚unbekannte' Wahrscheinlichkeit $p = 10^{-4}$ auf Basis einer Stichprobe vom Umfang $n = 10.000$.

Schätzwert	0	0,0001	0,0002	0,0003	0,0004
Wahrscheinlichkeit %	36,8	36,8	18,4	6,3	1,8

3.6 Axiomatik und ihre Rechtfertigung für eine Wahrscheinlichkeitskonzeption

Wahrscheinlichkeiten werden oft auf relative Häufigkeiten reduziert. Dabei wird die Möglichkeit einer Interpretation mit der ausschließlichen Gültigkeit dieser Interpretation verwechselt. Der Fokus auf relative Häufigkeiten und deren Entwicklung hat negative Auswirkungen auf das intuitive Verständnis (Borovcnik, 2023). Anstatt zu klären, dass die Muster der Entwicklung relativer Häufigkeiten keine Rolle spielen, werden diese Muster geradezu als wichtig erachtet. Bedingte Wahrscheinlichkeiten werden entsprechend zu Nebensachen degradiert und können vernachlässigt werden.

Es existiert auch ein Axiomensystem für Präferenzen, in welchem der Begriff bedingte Wahrscheinlichkeit eine Schlüsselrolle übernimmt. In der Axiomatik von Kolmogorov dagegen spielen bedingte Wahrscheinlichkeiten eine untergeordnete Rolle und werden erst in einer entsprechenden Definition aufgefangen (Borovcnik, 2021). Die wesentlichen Kenngrößen statistischer Entscheidungen basieren auf bedingten Wahrscheinlichkeiten (Borovcnik, 2023). Das zeigt, dass der weitgehend ausschließliche Bezug auf Kolmogorov im Schulunterricht und an der Universität wenig zur Klärung der Begriffe beiträgt.

3.7 Risiko und Entscheidungen

Sprache ist verräterisch: „*Das Risiko gehe ich nicht ein.*" Hier geht es darum, den schlimmsten Fall, das maximal Unangenehme, abzuwehren. Beispiele dafür sind die Maßnahmen gegen BSE in den 2000ern und die jüngeren COVID-Maßnahmen. Oder: „*No risk–no fun*." Das Schicksal herauszufordern ist eine große Triebfeder. Schon Kahneman und Tversky (1979) zeigen, dass hoher Einfluss (Folgekosten, Nutzen) die Wahrnehmung kleiner Wahrscheinlichkeiten stark beeinflusst (für eine Re-Analyse siehe Borovcnik, 2016b). Anstatt in einem Szenario relative Risiken gegeneinander abzuwägen („weigh the evidence") und Entscheidungskriterien offenzulegen, wird irrational entschieden. ‚Risiko' wird auch verwendet, um Wahrscheinlichkeiten von Fehlentscheidungen (bei statistischen Tests) zu beschreiben. Eine Elementarisierung statistischer Tests durch das Weglassen von Zusammenhängen (Typ I und II) sowie ein Transfer auf die Datenebene – durch Simulation – zerstört die Begriffe (Borovcnik, 2021).

3.8 Reichweite des Ansatzes und seiner Methoden

Stochastische Modelle können immer eingesetzt werden, solange es sich um einen normalen ‚Geschäftsverlauf' handelt. Bei Brüchen in Entwicklungen versagen diese Methoden. Hier hilft mehr Wissen über den Kontext oder, noch besser, Insiderwissen.

- *Korrelation oder Assoziation hat (etwas) mit Kausalität zu tun:* Dadurch erspart man sich oft substanzwissenschaftliche Forschung über Wirkmechanismen (Borovcnik, 2016a). Dabei sollten signifikante Korrelationen nur als Anlass gesehen werden, um ernsthaft über sachliche Zusammenhänge zu forschen.
- *ABC-Ansatz:* Die wichtigsten Dinge zuerst erledigen. Statistik ist auf Level C.
- *Experten sind der Wahrheit verpflichtet:* Experten haben natürlich Interessen und sind oft anders von Entscheidungen betroffen (z.B. sind sie juristisch haftbar).

3.9 Objektiver Stellenwert von Modellen

- *Es gibt nicht das eine, ‚beste' Modell:* Oft werden stochastische Modelle parallel durchgerechnet, um mehrere plausible Szenarien für die Entwicklung durchzuspielen. Das ist auch der Standardfall in der Beurteilenden Statistik.
- *Man kann Modelle wissenschaftlich (objektiv) evaluieren:* Dafür sind statistische Tests da. Doch die Rationalitätskriterien versagen für solche Tests, was sie zu bloßen Plausibilitätsprüfungen degradiert. Zielführender sind fundamentale Ideen, welche die Modelle beschreiben und die als Schlüssel zur Anwendung oder Verwerfung bestimmter Modelle für spezielle Situationen dienen.

4. Knackpunkte in der Wahrscheinlichkeitsrechnung

4.1 Die Rolle der Gleichwahrscheinlichkeit

- Gleichwahrscheinlichkeit spielt eine zentrale Rolle.
- Sie ist nur im Zusammenhang mit der Interpretation als relative Häufigkeit und als Wahrscheinlichkeitsverhältnis (subjektiver Vertrauensgrad) sinnvoll. Gleichwahrscheinlichkeit kann nur über relative Häufigkeiten erfasst werden – Wahrscheinlichkeit über längere Zeiträume hinweg.
- Gleichwahrscheinlichkeit kann zur Kalibrierung des subjektiven Vertrauensgrades in eine unsichere Situation dienen – Wahrscheinlichkeit für Einzelentscheidungen.
- Gleichwahrscheinlichkeit kann auch zur Hypothesengenerierung genutzt werden – Varga (1983) untersucht bereits im Anfangsunterricht, vor Einführung des Wahrscheinlichkeitsbegriffs, ob ein Münzwurfprotokoll echt oder ausgedacht ist (und führt intuitiv zum Signifikanztest).

Es ist wichtig, einen möglichst allgemeinen Wahrscheinlichkeitsbegriff zu entwickeln und den weit verbreiteten *Equiprobability Bias* (Lecoutre, 1992) kritisch zu hinterfragen.

4.2 Der Nutzen von Wahrscheinlichkeiten

Was bedeuten Aussagen wie P(E) = ½, P(E) = 1/6? Wozu dienen solche Aussagen?

Man kann damit Wetteinsätze bestimmen: 1:1 bzw. 5:1 (netto). In Serien kann jedoch alles passieren, und das Feedback ist nur eingeschränkt nutzbar.

- Jemand, der diese Aussagen in vielen Situationen anwenden kann (Gesetz der großen Zahlen), kann das Risiko weitgehend minimieren.
- Jemand, der nur eine Einzelentscheidung trifft, trägt das volle Risiko.

Verschiedene Stakeholder müssen ganz unterschiedliche Entscheidungskriterien anwenden (Borovcnik, 2019). Zwei Beispiele sollen dies verdeutlichen:

Entscheidung über die Auflagenhöhe: Optimierung einer Entscheidung

Die Nachfrage N nach einer Zeitschrift sei durch das Szenario in Tabelle 2 modelliert, die Kosten hängen affin-linear von der Auflagenhöhe ab: $C(a_j) = 2000 + (a_j - 1000)\cdot 0{,}2$. Der Preis pro Heft beträgt 1,60. Man kann 1000, 2000, ... , 5000 drucken. Wie viele sollte man drucken? (Vgl. Borovcnik, 2016b).

Tabelle 2: Wahrscheinlichkeitsverteilung der Nachfrage N. Bestimmung der Verteilung des Absatzes bei Option a_2 =2000.

$N = n_i$	1000	2000	3000	4000	5000	Profit bei a_2	
$p(n_i)$	0,40	0,30	0,20	0,06	0,04	Erwartung	max
Absatz bei a_2	1000	2000	2000	2000	2000	360	–600

Der Absatz ist eine Zufallsvariable. Bei Option a_2 ergibt sich somit ein erwarteter Profit von 360 bei einem Risiko, 600 € zu verlieren. Berechnet man diese Kennziffern für alle Optionen, so ergeben sich die Werte in Tabelle 3. Der Vergleich der Optionen erfolgt anhand des erwarteten Profits (Nutzen). Das Optimum liegt bei einer Auflage von 3000. Die Entscheidung hängt jedoch vom verwendeten Kriterium ab. Der maximale Verlust führt – nicht nur hier – zu unsinnigen Entscheidungen, nämlich gar nichts zu drucken. Man könnte auch den Nutzen anstelle des Geldes als Maßstab heranziehen.

Tabelle 3: Vergleichen der zukünftigen Entwicklung hinsichtlich des Profits – die Entscheidung hängt vom Kriterium ab.

Option	1000	2000	3000	4000	5000
Erwarteter Profit	–400	360	640	600	464
Max Verlust	–400	–600	–800	–1000	–1200

Versicherungskontrakt

Bei einer Wette auf ein Ereignis E stehen die Chancen $p : q$. Diese so genannten Odds sind das Verhältnis der Wahrscheinlichkeit des Eintretens zum Ausbleiben von E. Es gilt: $P(E) = p / (p + q)$. Die Wettquoten ergeben sich als die umgekehrten Odds, die Nettoeinsätze der Wettpartner verhalten sich wie $q : p$. Angenommen, Sie möchten eine Kaskoversicherung für Ihr neues Auto abschließen. Der Einfachheit halber betrachten wir nur die Fälle unfallfrei und Totalschaden (mit 20.000 € Schaden). Schätzen Sie ‚Ihre' Wahrscheinlichkeiten für ein Jahr und bestimmen Sie daraus den Wert der Entscheidung, keine Versicherung abzuschließen. Vergleichen Sie dies mit den Kosten der Versicherung (1.000 € Prämie). Sollten Sie die Versicherung abschließen? (Vgl. Borovcnik, 2016b). Bestimmen Sie zusätzlich jene Odds für einen Totalschaden, bei denen die Entscheidung zwischen Versicherung ja bzw. nein ‚kippt'.

Im Nachhinein ist jeder klüger: kein Unfall, keine Versicherung! Wenn man die maximalen Kosten minimiert, ist Versicherung (Option A_1) die bessere Wahl. Ist man bereit, die zukünftigen Ereignisse T_1 (kein Unfall) und T_2 (Totalschaden) zu gewichten, etwa durch relative Gewichte von 39 : 1, dann sinken die ‚erwarteten' Kosten für keine Versicherung (Option A_2) auf 500 €. Die Entscheidung hängt von den Gewichten ab. Welche Gewichte? Da es schwierig sein mag, diese Gewichte genauer anzugeben, könnte man auf den ‚Break-even-Point' ausweichen; der liegt bei Gewichten von 19 zu 1. Bei diesen Gewichten sind beide Optionen gleichwertig.

4.3 Kleine Wahrscheinlichkeiten

Kleine Wahrscheinlichkeiten sind heute unverzichtbar, bereiten jedoch große Probleme (Borovcnik, 2022). Noch in den 1950er Jahren schlugen namhafte Mathematiker (z.B. E. Borel) vor, Wahrscheinlichkeiten unterhalb eines bestimmten Schwellenwerts als null zu betrachten – historisch lief dies unter dem Begriff der „moralischen Wahrscheinlichkeit". Kleine Wahrscheinlichkeiten können empirisch kaum validiert werden, da es einfach keine verlässlichen Informationen darüber gibt. Wenn überhaupt, kann man sie in Form von Szenario-Größen verwenden – auf der Basis von „was wäre, wenn ...?". Solche Szenarien sind jedoch nur im *Vergleich* interpretierbar, nicht in ihren absoluten Werten. So kleine Wahrscheinlichkeiten sind schwer von den Folgekosten zu unterscheiden und führen häufig zu völligen Fehleinschätzungen.

Beispiel: Prävalenz (Verbreitungsgrad, Anteil der Population, der davon betroffen ist) von BSE oder von COVID. Die Prävalenz ist nicht nur orts-, sondern auch zeitabhängig, je nach Verlauf des Infektionsgeschehens.

4.4 Information in Wahrscheinlichkeitsverteilungen

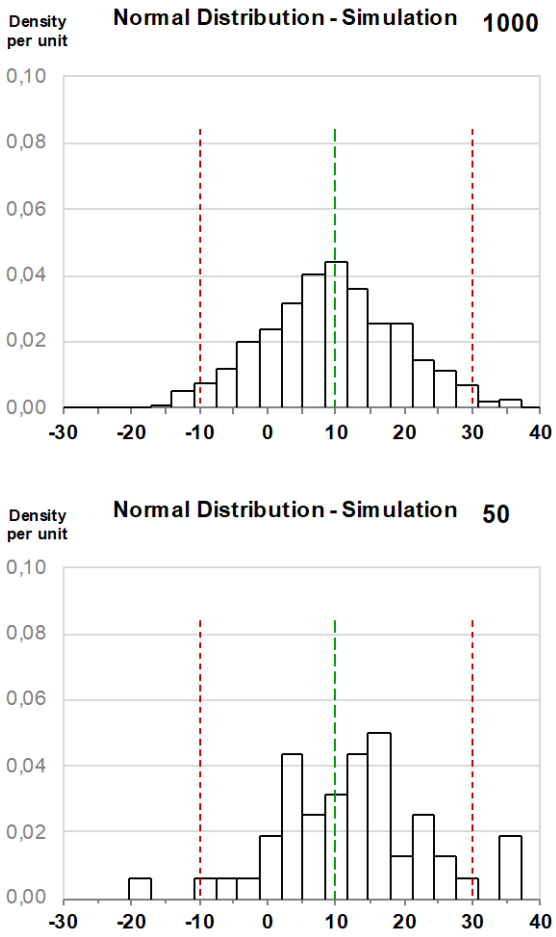
Daten zu Wahrscheinlichkeitsverteilungen schwanken weit mehr, als man sich vorstellen kann. Es ist instruktiv, sich bestimmte Verteilungen simulieren zu lassen, um ein Gefühl für diese Schwankungen zu bekommen. Dies macht man am besten im statistischen Labor; ansonsten neigt man allzu sehr dazu, kontextabhängige Erklärungen für diese Schwankungen zu suchen. Weder eine statistische noch eine probabilistische Herangehensweise ist die folgende, welche sehr populär ist. Danach werden Verteilungen nach kleinsten Anomalien und Besonderheiten untersucht. Wichtig ist es jedoch, den *Typ* der Verteilung einzuschätzen: Eingipfelig, symmetrisch, schief etc. Die Information, die in Wahrscheinlichkeitsverteilungen enthalten ist, hängt entscheidend von folgenden Faktoren ab:

- *Typ der Verteilung.* Die Einschätzung, ob die Verteilung eingipfelig, symmetrisch oder schief ist, ist von elementarer Bedeutung. Die Bewertung, ob einzelne Werte ‚extrem' oder ‚normal' für die Verteilung sind, hängt entscheidend davon ab.
- Reduktion auf Erwartungswert und Standardabweichung sowie z-Werte. Für buchhalterische Zwecke (Gesetz der großen Zahlen) ist diese Reduktion besonders sinnvoll, wenn die Verteilung ausgeprägt eingipfelig und im Wesentlichen symmetrisch ist. Die sogenannten *z-Werte* spiegeln dann ein absolutes Ranking wider.
- Idee hinter der Verteilung, welche zur Modellierung hilfreich ist. Einige Beispiele:
  - Gleichverteilung – Fairness, Symmetrie
  - Binomialverteilung – Unabhängige Wiederholung desselben 0-1 Experiments
  - Poisson- und Exponentialverteilung als Zählen von bzw. Warten auf Ereignisse im völlig zufälligen Entstehen von Ereignissen.
  - Normalverteilung als Folge der *Elementarfehlerhypothese* (ZGS)**.** Diese rechtfertigt die Normalverteilung als Näherung für fast alle Schätzungen von Parametern aus Stichproben (= Summen).

Fragen, die schon in Richtung Signifikanztests gehen, schließen sich an die Untersuchung, wo ‚extreme' und wo ‚normale' Werte der Verteilung liegen. Wie groß sind die Zufallsschwankungen. Wann gehört ein Wert noch zu dieser Verteilung oder kommt aus einer anderen ‚Quelle'? Wo liegen die Trennpunkte, welche die bestehende Verteilung von anderen abgrenzt? Was haben diese Schwellenwerte zu bedeuten? – Modellkritik: Kann die Verteilung als Modell für ein zu untersuchendes Phänomen dienen. Welche Konsequenzen ergeben sich, wenn man Entscheidungen auf Basis dieser Verteilung trifft und die Schwellenwerte zur Abgrenzung von anderen Verteilungen so und nicht anders festlegt?

Kleine Risiken, die Schwellenwerte von 2 Standardabweichungen zu überschreiten

Bei nur 50 simulierten Daten schwankt der Prozentsatz innerhalb des Schwellenintervalls erheblich. Werte zwischen 85 und 99% sind nicht ungewöhnlich. Die Verteilung sieht zerrissen aus und erinnert kaum an eine Normalverteilung; oft fehlt sogar ein ausgeprägter Gipfel. Wenn σ klein ist, kann ein einziges Datum bereits die Lage der Verteilung (den Erwartungswert) anzeigen! Es besteht nur ein geringes Risiko, dass die Abweichung größer als zwei Standardabweichungen ist. Für Mittelwerte bietet die Normalverteilung in der Regel eine ausreichende Annäherung an die tatsächliche Verteilung.



Man beachte die erhebliche Variabilität in der Form der Verteilung der simulierten Daten, insbesondere bei kleinen Stichproben. Es handelt sich um Daten aus einem statistischen Labor, die keine Störfaktoren enthalten, wie sie in der Realität auftreten und das Ergebnis wesentlich variabler und unübersichtlicher gestalten könnten.

Stichproben-umfang	% Daten innerhalb von ± 2*s*
1000	94,9%
50	88,0%

Man erkennt: Idealerweise, d.h. bei ausreichender Datenmenge, bleibt der Prozentsatz der Daten innerhalb von zwei Standardabweichungen des Erwartungswerts stabil bei 95%, unabhängig von den Parameterwerten (die in der App über Regler verändert werden können).

Abb. 1: Die Risiken für extreme Werte schwanken nur geringfügig (um etwa 5%) im Vergleich zur gezeigten Verteilung. Die Schwellenwerte sind bei größerer Streuung weiter voneinander entfern und verschieben sich mit dem Erwartungswert.

5. Knackpunkte in der beurteilenden Statistik

5.1 Die Stichprobenverteilung – für den Mittelwert

Warum verwendet man Mittelwerte (oder andere Parameter) aus einer Stichprobe als Stellvertreter für die gesamte Verteilung? Warum ist die Normalverteilung so wichtig? Wichtiger als die Beschreibung realer stochastischer Phänomene X durch die Normalverteilung ist ihre Verwendung als approximative Verteilung für eine *künstliche* Zufallsvariable. Wir können die Verteilung von X beobachten und erhalten die Daten einer Stichprobe: $x_1, x_2, \ldots, x_n$. Daraus berechnen wir den Mittelwert $\bar{x}$. Davon haben wir einen *einzigen* Wert; dieser repräsentiert die Stichprobe. Da der Mittelwert μ der Verteilung von X in der Regel unbekannt ist, müssen wir eine *künstliche* Vergleichsbasis schaffen.

Bestimmung der Stichprobenverteilung

Durch mathematische Methoden oder Simulationen im statistischen Labor untersuchen wir, wie die Verteilung des realen Phänomens X mit der Verteilung des Mittelwerts $\overline{X}$ einer Stichprobe zusammenhängt. Zwei wesentliche Eigenschaften treten dabei hervor:

a) Die Verteilung der Mittelwerte $\overline{X}$ lässt sich annähernd durch eine Normalverteilung beschreiben (Zentraler Grenzwertsatz). Es genügt, ihren Mittelwert sowie ihre Standardabweichung zu kennen.

b) Die Parameter der Verteilung der Mittelwerte $\overline{X}$ hängen vom Mittelwert und der Standardabweichung der Ausgangsverteilung von X so ab: $E(\overline{X}) = \mu$ und $\sigma(\overline{X}) = \sigma / \sqrt{n}$.

Man beachte: Während wir von X über die Stichprobe immer mehrere Werte beobachten, haben wir von $\overline{X}$ immer nur einen *einzigen* Wert. Damit erhalten wir eine artifizielle Verteilung, mit der wir den beobachteten Wert von $\overline{X}$ vergleichen können. Ganz im Sinne der Schwellenwertüberlegung.

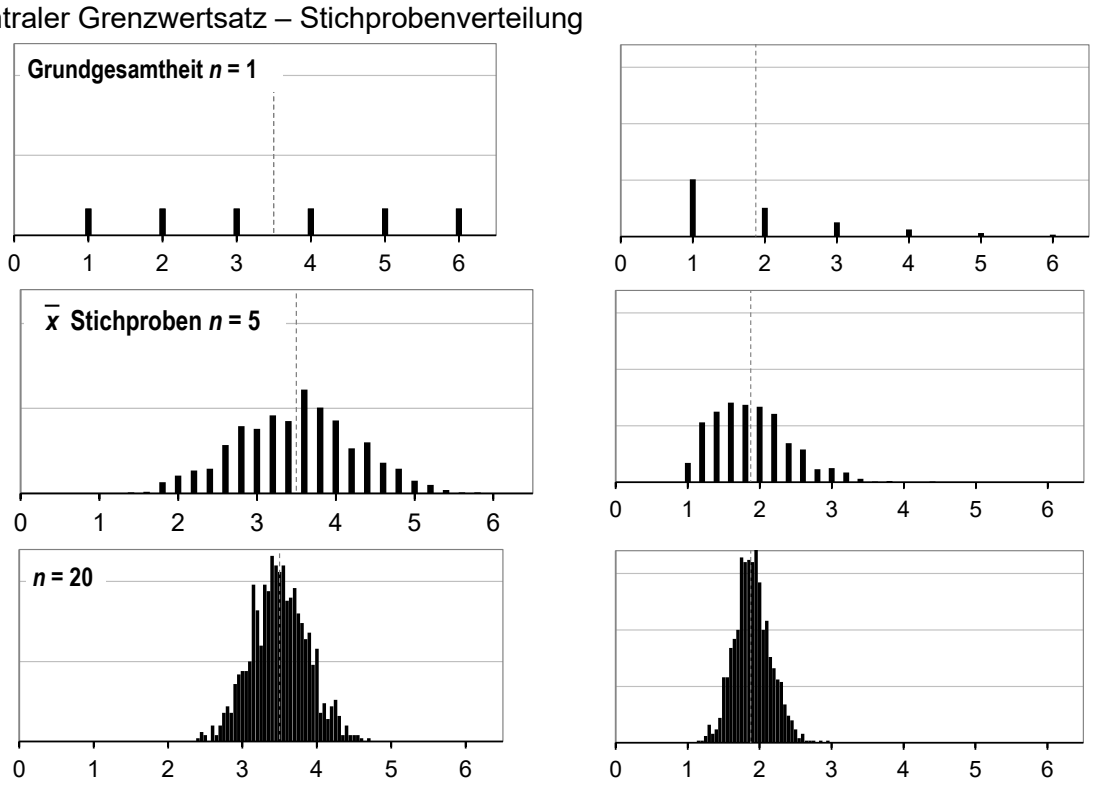
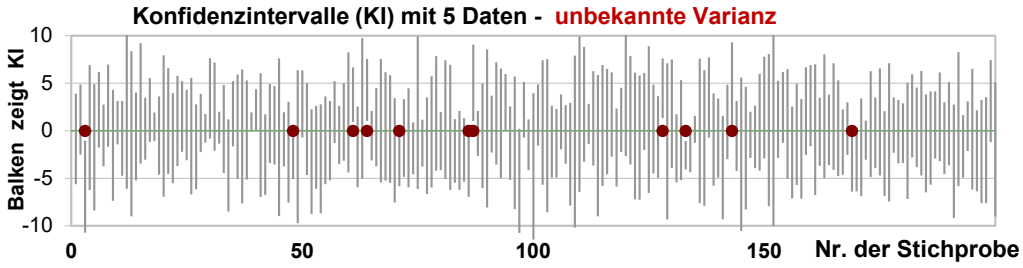


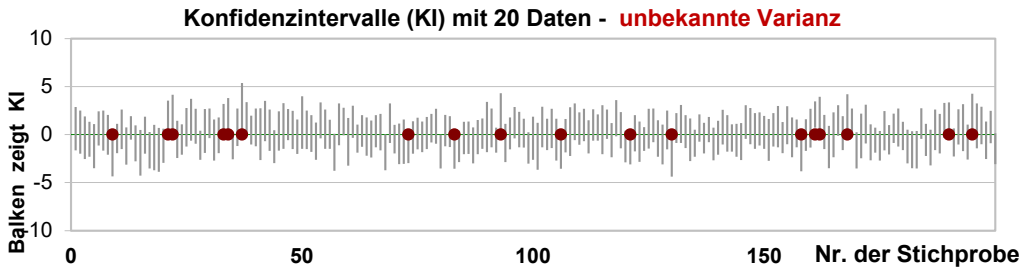
Abb. 2: Grundgesamtheit und Verteilung der Stichprobenmittelwerte bei Stichprobenumfängen *n* = 5 und *n* = 20. Links eine Gleichverteilung, rechts eine J-förmige Verteilung. Die Verteilung der Mittelwerte normalisiert sich sehr schnell. Dies verdeutlicht die Bedeutung der Normalverteilung, insbesondere für die Stichprobenverteilung von Parameterschätzungen.

5.2 Konfidenzintervalle

Bei der Analyse einer Verteilung betrachten wir zunächst alle möglichen Hypothesen. Diese werden durch Schwellenwertüberlegungen gefiltert, sodass nur einige wenige plausible Hypothesen übrigbleiben. Grundsätzlich gibt es zwei Methoden zur Analyse: Konfidenzintervalle und statistische Hypothesentests (siehe Abschnitte 5.3 und 5.4).

Ein *Konfidenzintervall* umfasst alle Werte von μ, welche den beobachteten Wert von $\bar{x}$ innerhalb der Schwellenwerte haben. Man sagt, dass eine solche Beobachtung mit dem Modellwert μ kompatibel ist. Ein Konfidenzintervall ist einfach ein Intervall kompatibler Werte. Es wird klar, dass das Risiko des Überschreitens des Schwellenwerts auch noch beachtet werden muss. Die Gegenwahrscheinlichkeit nennt man Überdeckungswahrscheinlichkeit. Ein grundsätzliches Missverständnis ist die Annahme, dass ein realisiertes Konfidenzintervall den Parameter μ mit einer bestimmten Wahrscheinlichkeit abdeckt (siehe Abbildung 3). Die Darstellung zeigt, wie stark sich die Überdeckungswahrscheinlichkeit von Intervallen, die sich auf kurze, mittlere und lange Intervalle bezieht, unterscheiden (dabei wird die Varianz des Merkmals als unbekannt angenommen).





	Prozentsatz überdeckender Intervalle			
	ALLE	KURZE	MITTLERE	LANGE
5 Daten	94,5%	88,1%	95,5%	100,0%
20 Daten	91,0%	86,6%	90,9%	95,5%

Abb. 3: Die Überdeckungswahrscheinlichkeit trifft nicht auf das konkrete Intervall zu. Die Simulation von 200 Stichproben mit 5 (oben) und 20 Daten (unten) zeigt die realisierten Konfidenzintervalle, die durch Balken gekennzeichnet sind. Konfidenzintervalle, die den (hier bekannten) Erwartungswert nicht abdecken, sind durch einen Punkt markiert.

5.3 Hypothesentests – Filtern mit Vorbedingungen

Anstelle der künstlichen Stichprobenverteilung der Mittelwerte betrachten wir die tatsächliche Verteilung des Phänomens. In diesem Fall geht es um die Unterscheidung zwischen zwei Verteilungen – Gesunde und Kranke – anhand eines Diagnosemerkmals. Ziel ist es, einen geeigneten Trennpunkt zu finden. Jeder Trennpunkt hat in diagnostischer Hinsicht zwei gegensätzliche Fehlerraten (siehe Abbildung 4). Diese beziehen sich auf zwei verschiedene Szenarien, die gegeneinander abgewogen werden müssen: i. Gesunde werden als Kranke diagnostiziert. ii. Kranke werden als gesund eingestuft. Anstelle von bedingten Wahrscheinlichkeiten für Alpha- und Beta-Fehler arbeiten wir mit den konkreten relativen Häufigkeiten für Fehldiagnosen in zwei unterschiedlichen Szenarien (Vgl. Borovcnik, 2021).

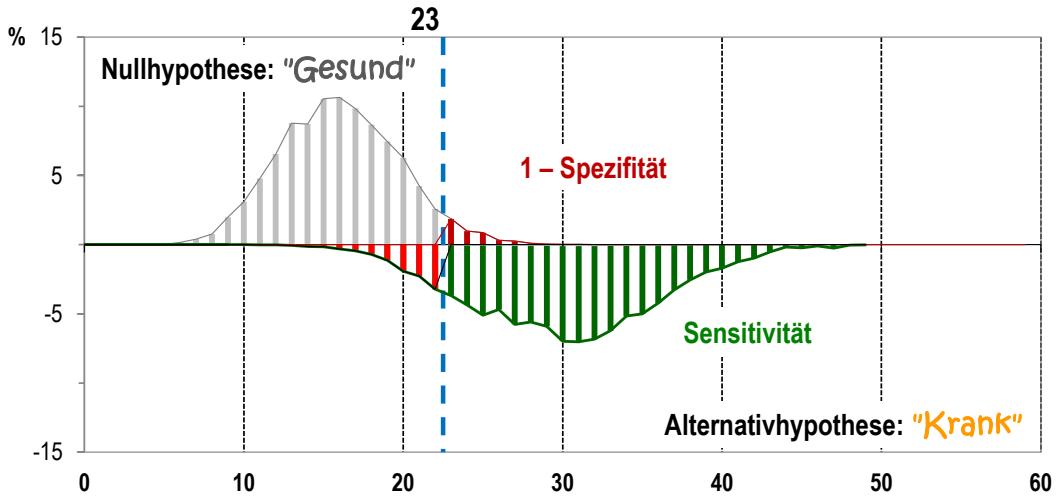


Abb. 4: Vorübungen zu Hypothesentests – Trennung zweier Szenarien. Die simulierten Daten eines Diagnosemerkmals für die Gruppen von Gesunden und Kranken zeigen, dass Kranke tendenziell höhere Werte aufweisen. Der Balken ‚trennt' Gesunde von Kranken. Personen mit einem Wert darüber werden als krank eingestuft, darunter als gesund.

5.4 Hypothesentests – ein Problemkind der Wissenschaften

Die Festlegung eines Fehlers führt zu einer asymmetrischen Problematik und vernachlässigt den jeweils anderen Fehler. Die mathematische Vorgehensweise bei einem Kontinuum von Szenarien besteht darin, den sogenannten Alpha-Fehler festzulegen und den Beta-Fehler zu optimieren. Beim reinen Signifikanztest – der Nullhypothese ohne Spezifikation einer Alternativhypothese – entfällt das Optimalitätsargument vollständig. Leider kann man in der Praxis weder auf statistische Tests (zugunsten von Konfidenzintervallen) verzichten, noch kommt man ohne reine Signifikanztests aus. Die Angabe von p-Werten ist besonders bei reinen Signifikanztests verwirrend und führt häufig zu Fehlinterpretationen bezüglich der Gültigkeit der Nullhypothese (Hubbard & Bayarri, 2003; ASA, 2016. In wissenschaftlichen Zeitschriften wird die Praxis der reinen Signifikanztests und p-Werte daher etwa alle zehn Jahre erneut hinterfragt.

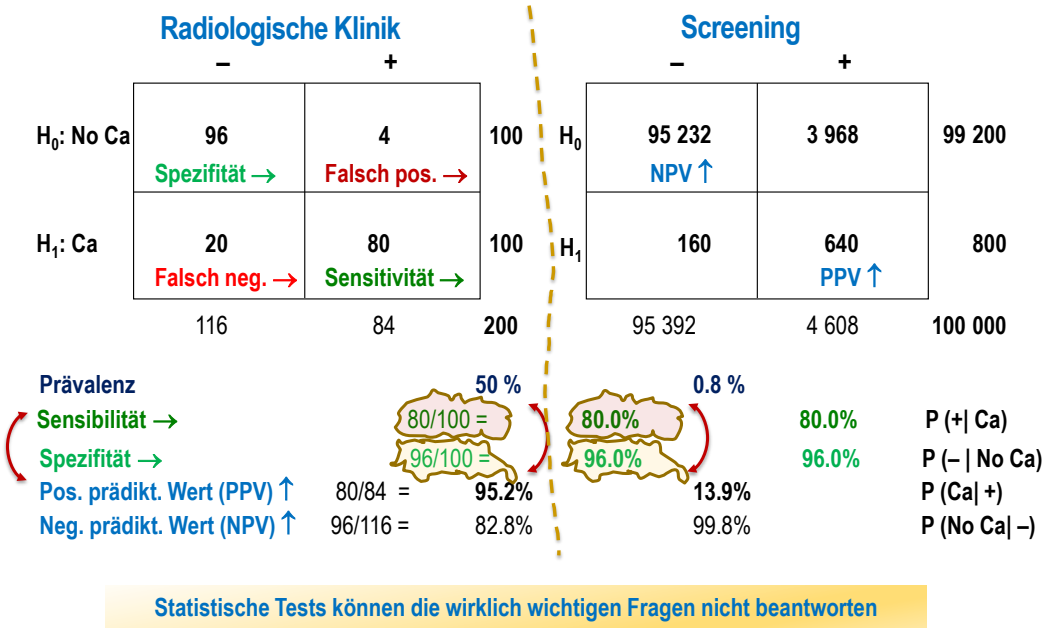


Abb. 5: Statistische Tests durch den Bayes-Ansatz verstehen. Fehlendes Bindeglied zur Beurteilung der Diagnosefehler: Die Prävalenz. Verglichen werden zwei Szenarien zur Diagnose von Krebs in der radiologischen Klinik (links) sowie im Screening (rechts). Obwohl Sensitivität und Spezifität der Diagnose gleichbleiben, unterscheiden sich die prädiktiven Werte hinsichtlich der Vorhersagekraft erheblich (95,2% tatsächlich krank bei positiver Diagnose in der Klinik, aber nur 13,9% im Screening).

Der Kontext der Medizin veranschaulicht, was bei klassischen Tests fehlt: die a-priori-Wahrscheinlichkeit der Nullhypothese. Dieser Kontext bietet viel Potential, statistisches Testen verstehen zu lernen (Abbildung 5; siehe Borovcnik, 2022). Die Qualitätsindizes sind in beiden Situationen identisch: Sensitivität = 80%, Spezifität = 96%. Dennoch variiert die Qualität der Entscheidungen erheblich: In der Klinik mag die Diagnose unterstützend sein, im Screening ist sie jedoch nahezu wertlos. Die Diskrepanz wird umso größer, je kleiner die Prävalenz ist. A-posteriori-Wahrscheinlichkeiten haben ohne Bezug auf a-priori-Wahrscheinlichkeiten keinen Sinn und entbehren einer Häufigkeitsdeutung, die über ein eng gefasstes Szenario hinausgeht. Diese Häufigkeitsdeutung ist für das Individuum nutzlos; sie dient lediglich als Referenzpunkt für das System und als Argumentationshilfe zur Rechtfertigung der Diagnostik.

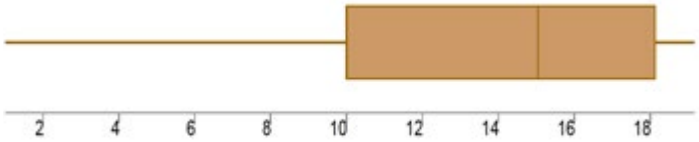
6. Knackpunkte in der beschreibenden Statistik

6.1 Multiple-Choice-Test zur Studieneingangs- und Orientierungsphase

Bei einem Multiple-Choice-Test zur Studieneingangs- und Orientierungsphase im Fach ‚Technische Mathematik' nahmen 24 Anfänger teil, 17 davon beim ersten Prüfungstermin. Es handelt sich um eine Knock-out-Prüfung: Man hat zwar die üblichen Wiederholungsantritte, aber man muss diese Prüfung erfolgreich abschließen, um weiterstudieren zu können. Die Vorlesung umfasst unter anderem 9 Stunden Statistik, begleitet von einem Skriptum mit 30 Seiten, das sich auf statistische Methoden konzentriert. Die behandelten Themen umfassen die Beschreibung univariater Daten (Kennziffern, Boxplot, Stabdiagramm, Histogramm, Lorenzkurven) sowie die Beschreibung bivariater Daten und die Vorgehensweise bei Regression und Korrelation. Der Test besteht aus Fragen zur Deskription univariater Daten (5 Items), Lorenzkurven (3) sowie Regression und Korrelation (2). Im Folgenden werden einige Testfragen diskutiert, es wird aufgezeigt, welche Antwortkategorien zur Auswahl standen, wie häufig sie gewählt wurden und welche fundierten Begründungen für die richtige Antwort gegeben werden könnten. Die Auswahl der Items und die Art der Fragestellung ist Indiz dafür, was dem Autor als Knackpunkte in der beschreibenden Statistik vorschwebt: Eine fundamentale Literalität, flüssiges Lesen von Daten, Kennziffern und Darstellungen, sowie Herstellen eines Bezugs zum Kontext.

#3. Boxplot.

Man betrachte den folgenden Boxplot.



Welche der folgenden Aussagen ist wahr?

I. Die Verteilung ist rechtsschief (linkssteil).
II. Der IQR ist ca. 8.
III. Der Median ist ungefähr 10.

☐ Lediglich I.
☐ Nur II.
☐ Nur III.
☐ I und III.
☐ II und III.

4. Mittelwert und Standardabweichung. Mittelwert und Standardabweichung von x-Daten sind gegeben durch: $\bar{x} = 100\,013{,}4$, $s = 10{,}23$. Neue y-Daten werden durch eine affin-lineare Transformation der x-Daten wie folgt festgelegt: $y = \frac{x - 100\,000}{10}$. Welche der folgenden Aussagen ist wahr?

I. Man kann aus den Angaben den Mittelwert der y-Daten ausrechnen, nicht aber deren Standardabweichung.
II. Aus den Angaben kann man sowohl Mittelwert als auch Standardabweichung der y-Daten berechnen.
III. Es gilt: $\bar{y} = 13{,}4$ und $s_Y = 0{,}1023$.

☐ Lediglich I.
☐ Lediglich II.
☐ Lediglich III.
☐ II und III.

5. Graphische Darstellungen von Daten

Daten zum Alter aus einer Untersuchung (rechts). Ist die Darstellung der Verteilung des Alters (unten) korrekt?

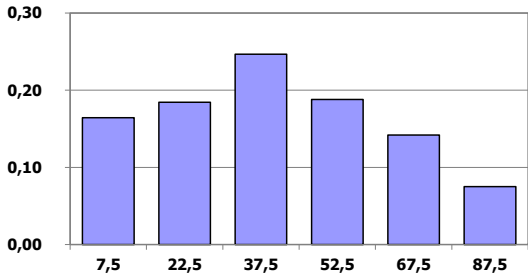
☐ Ja, ist korrekt.
☐ Nein, ist nicht korrekt.
☐ Ja, aber man sollte die Abstände zwischen

Altersbereich von	bis		Mitten	Anzahl n_i	Anteil h_i
0	bis inkl.	15	7,5	1.333.505	0,16
15	- " -	30	22,5	1.495.740	0,18
30	- " -	45	37,5	2.002.259	0,25
45	- " -	60	52,5	1.526.110	0,19
60	- " -	75	67,5	1.151.122	0,14
75	- " -	100	87,5	609.018	0,08
				8.117.754	1,000

den Säulen vermeiden.

☐ Ja, aber man sollte die Klassenmitten so wählen, dass schönere Zahlen herauskommen, damit die Graphik besser lesbar ist.

b) Histogramm der Altersverteilung



6. Mittelwerte

I. Der Mittelwert ist der beste Wert, mit dem man eine Liste von Daten beschreibt.

II. Eine Gruppe von drei Personen hat folgende IQ-Werte (Intelligenzquotient): 70, 100, 130. Im Mittel haben sie einen IQ von 100.

III. Der Mittelwert von Daten $x_1, x_2, \ldots, x_n$ optimiert die Summe $\varphi(c) = \sum_{i=1}^{n} |x_i - c|$.

☐ Alle drei Aussagen treffen zu.

☐ Nur die Aussage III trifft zu.

☐ Nur die Aussagen II und III treffen zu.

7. Maße für die Lage einer Verteilung.

Aus einer großen Liste von Daten $x_1, x_2, \ldots, x_n$ hat man folgende Kennziffern berechnet: $\bar{x} = 18$, $x_{med} = 15$, $x_{\text{mod}} = 17$ und $s = 2$.

Was sagt das über die Verteilung der Daten aus?

☐ Die Verteilung ist linkssteil (rechtschief).

☐ Die Verteilung ist rechtsteil (linksschief).

☐ Wahrscheinlich ist bei der Berechnung ein Fehler passiert.

6.2 Antworten und Begründungen

Angegeben wird, wie häufig die Antwortkategorien angekreuzt wurden sowie, ob Begründungen richtig (r) oder falsch (f) waren oder gefehlt haben (↯). Ferner eine Auswahl an ‚Begründungen'. Richtige Begründungen treffen wesentliche Teile und können auch mit falschen Antworten verbunden sein.

3. Boxplot

4Wahrscheinlic	Lediglich I.		
12Wahrscheinli	Nur II.	r	**8**
0Wahrscheinlic	Nur III.	f	**7**
0Wahrscheinlic	I und III.	↯	**2**
1Wahrscheinlic	II und III.		

Median = 10 / 14. Rechtsschief wegen $x_{0,5} < x_{0,75}$. Linksschief → mehr Werte rechts.

4. Mittelwert und Standardabweichung

6	Lediglich I.		
6	Lediglich II.	r	7
0	Lediglich IIII.	f	5
2	II und III.	↯	5
1	Keine.		
1	Alles falsch.		
1	Nicht bearbeitet..		

s kann berechnet werden, wenn man zuvor alle y ermittelt, allein aus dieser Formel jedoch NICHT.

Alles falsch, s und $\bar{x}$ bleiben gleich.

5. Graphische Darstellungen

4	Ja, ist korrekt.	r	**5**
8	Nein, ist nicht korrekt.	r*	**6**
5	Ja, aber … sollte die Abstände …	f	**5**
0	Ja, aber … sollte ... Klassenmitten …	↯	**1**

Da Werte metrisch sind, benötigt man keine Abstände.

6. Mittelwerte

1	Alle drei Aussagen treffen zu.		
0	Nur Aussage III trifft zu.	r	**3**
14	Nur Aussagen II und III treffen zu.	f	**9**
1	Keine Aussage ist richtig.	↯	**5**
1	Nur II.		

Man könnte das Diagramm falsch verstehen.
Man kann aus … Grafik die Werte gut ablesen.
Abstände sind egal.
Mittelwert dient zur Veranschaulichung.
Mittelwert (70+100+130)/3 = 100.
Der Mittelwert passt nicht, wenn Ausreißer …

7. Maße für die Lage einer Verteilung. ...

2	Die Verteilung ist linkssteil.	r	9
5	Die Verteilung ist rechtssteil.	f	6
10	Wahrscheinlich ist … Fehler passiert.	√	2

Es befinden sich ~~mehr~~ höhere Daten im 2. Quartil → Steigung auf der rechten Seite.

Einige führen die Bedingung *mean* > *med* bzw. *mean* > *mod* an (was am Problem vorbeigeht).

Tab.5: Häufigkeiten der Antworten und Begründungen (richtig r, falsch f, fehlend √). Ferner eine Auswahl an ‚Begründungen'.

Vorschläge für passende Begründungen

#3 Der linke Teil der Box ist breiter als der rechte, daher ist die Verteilung linksschief; der Strich für den Median liegt bei 15, die Box hat eine Breite von 18−10 = 8.

#4 Zwischen den Kennziffern der Daten und der affin-linear transformierten Daten besteht ein direkter Zusammenhang; die Rechnung in III ist aber falsch.

#5 Merkmal ist metrisch, daher keine Abstände, Achse ist falsch unterteilt, die Höhen (anstelle der Flächen) stellen die r. H. dar.

#6 I hängt vom verwendeten Kriterium ab; II Daten sind *nicht* metrisch; III Mittelwert minimiert *quadrierte* (!) Abweichungen.

#7 Die angegebenen Werte verletzen die Faustregel *mean* ≤ *med* ≤ *mode* (oder in umgekehrter Reihenfolge). Mit Ausnahme von höchst artifiziellen Daten kann der Modus nicht strikt zwischen Mittel und Median liegen.

7. Zum Abschluss

7.1 Einschätzung der Verwendung von Stochastik in verschiedensten Disziplinen

Öffentlicher Umgang mit stochastischen Aussagen. Ein Prominenter äußert sich... (z. B. Angelina Jolie und die 80% Risiko). Ein Experte rät... (z. B. ab dem 30. Lebensjahr jährlich zur Mammographie, insbesondere bei familiärem Risiko). Am Ende steht ein Ranking von Belegen: Wie bestimmte Genkonstellationen den Stoffwechsel beeinflussen, wann und warum präkanzeröse Zellveränderungen in bösartiges Tumorwachstum übergehen. Aber was nützt das? Wir leben in einer ‚Evidence-based Society'.

Vorhersage wirtschaftlicher Entwicklungen. Mein verstorbener Mentor Sepp Gölles in Graz hat stets die ABC-Regel gepredigt und dabei die Statistik auf Stufe C eingeordnet: Zuerst die großen Ungereimtheiten mit maximaler Wirkung beseitigen, dann die Feinabstimmungen vornehmen, die vergleichsweise wenig Einfluss haben. Der Konflikt zwischen Lehman Brothers und Goldman Sachs und die unterschiedlichen Auswirkungen auf die USA und Europa ist A-Wissen. Ja, Insiderwissen ist oft wichtiger als Statistik, die ohnehin schwierig und unzuverlässig ist. Das ist keine Werbung für den langen Weg zur Statistik. Zudem ist der Austausch von Insiderwissen und das Schmieden von Intrigen menschlichen Bedürfnissen näher als die trockene Statistik.

Vorhersage von Erdbeben. Geologen arbeiten schon lange mit Stochastikern. Ich kenne seriöse Arbeiten dazu seit den 1970er Jahren. Sie müssten genügend Daten haben, wenn der Ansatz erfolgreich wäre. Dass immer wieder vage Vorhersagen von großen Beben gemacht werden (Kassandra hat am Ende immer recht), verbessert den Ruf der Profession nicht. Auf der anderen Seite wurden riesige Beben wie jene vor Sumatra oder Fukushima nicht vorhergesagt – nicht einmal annähernd. Das ist kein

Ruhmesblatt. Die Anklage gegen Erdbebenforscher in Emilia-Romagna durch italienische Gerichte zeigt die Frustration der Öffentlichkeit über Experten, die mit hohen Forschungsgeldern ausgestattet sind.

Wahlprognosen. Es geht nicht darum, die Feinheiten (oder Unsinnigkeiten) des US-amerikanischen Wahlsystems auszunutzen, um eine genaue Prognose zu erstellen. Der Prozentsatz global war als Gewinn für Hillary Clinton vorhergesagt. Entweder durften die Wahlforscher ihre Stärken nicht öffentlich zeigen, oder sie haben versagt. Es gibt immer wieder – auch bei uns – Fehlschläge bei Wahlprognosen.

Public Health. Viele Daten in diesem Bereich bestätigen das Sprichwort: ‚Garbage in – garbage out'.

7.2 Gedanken zu Unterricht und Überzeugungskraft

Stochastik scheint aus der Vogelperspektive betrachtet zu werden: Während sie auf organisatorischer Ebene für Entscheidungen und Bewertungen nützlich ist, bleibt der Einzelne außen vor.

Lotterie: Wie muss ich sie gestalten, damit sie interessant und ertragreich wird? Für den Einzelnen: „Ich spiele nicht, weil der Betreiber sowieso gewinnt." Dennoch spielen viele, und das nicht zu knapp. Was sind ihre Beweggründe und warum werden diese in der Stochastik nicht angemessen berücksichtigt?

Versicherung: Der Preis einer Polizze basiert auf statistischen Informationen und Faustregeln. Für die Versicherungsgesellschaft funktioniert das aufgrund des Gesetzes der großen Zahlen hervorragend. Für den Einzelnen hingegen bleiben seine individuellen Kriterien meist unberücksichtigt, sodass er dem Risiko voll ausgesetzt ist. Warum versichert sich der Einzelne dennoch?

Marktforschung: Die Klicks (insbesondere die Folge-Klicks) bei Amazon (wer bestellt was zusammen) und ein geeignetes Programm zur Aufzeichnung und Analyse ermöglichen es Amazon, situationsgerechte Angebote zu erstellen (viele Firmen machen das im Internet). Der Einzelne kann sich das weder leisten noch würde es ihm viel bringen – oder doch? Kleinere Anbieter haben den Nachteil, dass ihre Daten noch wenig Struktur aufweisen und daher weniger zuverlässig sind.

PISA: Die Betreiber haben eine Folgeindustrie aufgebaut, beginnend mit getesteten Aufgabenpaketen (die dann auch für zentrale Prüfungen genutzt werden) einschließlich Lösungen. Dass dabei so viele Fehler passieren (siehe Erich Neuwirth, o.D.), merkt man erst, wenn man sich jahrelang mit den Modellen beschäftigt. Die einzelnen Staaten können es sich nicht leisten, sich vom System abzukoppeln, und reformieren ständig ihre Schulsysteme aufgrund kaum interpretierbarer Rangunterschiede, bis sie ...

Wirtschafts- und Sozialstatistiken: Diese werden ständig angepasst, wenn sie nicht den ‚Erfordernissen' der ‚Betreiber' entsprechen. Dennoch dienen sie als Grundlage für tagespolitische Entscheidungen. Es gibt unendlich viele Tricks, die Zahlen so zu biegen, dass sie passen. Für viele organisatorische Planungen fehlen hingegen – oft aus Absicht – die notwendigen Statistiken. Vor etwa zehn Jahren hat man Lehrer mit verlockenden Angeboten in die Pension geschickt. Heute sucht man händeringend nach Lehrkräften und stellt viele auch ohne Prüfung ein. Dabei lässt sich der Bestand an Lehrkräften nach Alter, Ausfällen und Schülerzahlen gut abschätzen. Trotzdem schafft man es nicht, die Zu- und Abflüsse zu prognostizieren.

7.3 Die Knackpunkte in Stochastik

Einige Ansätze wurden hier vorgestellt, um die üblichen Konzepte besser verstehen und die darauf basierenden Methoden in ihrer Relevanz und ihren Einschränkungen beurteilen zu können. Ich glaube auch, dass die Grenzen der Methoden viel enger gesteckt sind als die Erwartungen an sie oder umgekehrt: Die Erwartungen an stochastische Methoden sind einfach zu hoch. Sie können nur so gut

sein wie die verwendeten Modelle. Diese erhält man nicht auf Knopfdruck, sondern nur durch sorgfältige Überlegung und tiefgehende stochastische Kenntnisse. Die Vorteile stochastischer Methoden sind für verschiedene Beteiligte unterschiedlich. Stakeholder, die die Methoden in ähnlichen Situationen häufiger einsetzen können, sind besser dran als diejenigen, die sie nur in Einzelfällen anwenden. Auch für diejenigen, die Einzelentscheidungen treffen, bieten stochastische Methoden eine unverzichtbare Orientierungshilfe, sei es, um sie zu nutzen oder besser zu verstehen, wenn sie damit konfrontiert werden.

Ja, die Physik kommt ohne Wahrscheinlichkeitstheorie nicht mehr aus – die gesamte Quantenmechanik basiert darauf. Aber das ist für die Sekundarstufe wohl etwas zu komplex. Auch der Finanzmarkt, speziell die Derivate, kommt ohne Stochastik nicht aus. Desgleichen die Versicherungsbranche. Künstliche Intelligenz baut Algorithmen häufig auf Wahrscheinlichkeiten auf. Auch hier scheint wenig Schulrelevantes vorhanden. Hier verschiebt sich die Frage nach dem, was man mit stochastischen Methoden besser machen kann als ohne, hin zur Feststellung, dass man ohne Stochastik gar nichts machen kann.

Das geringste Argument für die Stochastik ist, dass man beim ‚Lügen' aus dem Spiel ist, wenn man zu wenige einschlägige Kenntnisse hat. Es gilt die Knackpunkte zu finden, die den Zugang zur Stochastik als eine von mehreren Methoden rechtfertigen. Da haben wir den Unterricht noch einen langen Weg vor uns. Grundsätzliches Ziel der Stochastik-Ausbildung sollte sein zu klären, worin der eigene Ansatz besteht (Voraussetzungen, Zielkriterien), und ihn mit alternativen Ansätzen zu vergleichen.

Was die Prioritäten in den Konzepten anbelangt, die man im Unterricht behandelt, so ist in der beschreibenden Statistik eine Aufwertung einer allgemeinen Literalität anzustreben, das betrifft insbesondere ein flexibles Verständnis der Kennziffern und ein flüssiges Lesen von Diagrammen, alles immer auf den Kontext bezogen. Und, es geht fließend in Risikobetrachtungen, Wahrscheinlichkeitsüberlegungen und informelle statistische Beurteilung der Ergebnisse über (siehe etwa Borovcnik & Ossimitz,1987, oder Nordmeier, 1989). Hinsichtlich der Arbeitsform ist – nicht nur die beschreibende Statistik – eine Rückbesinnung auf projektorientierten Unterricht kein falscher Ansatz (siehe auch Borovcnik & Terán, 2023).

Der Szenario-Gedanke ist für die angemessene Modellierung stochastischer Situationen und insbesondere die Bewertung und Interpretation der Ergebnisse wesentlich (Borovcnik & Kapadia, 2011). Das zeigt auch, dass ein enger frequentistischer Rahmen für die Wahrscheinlichkeit eine Sackgasse darstellt. Was die Wahrscheinlichkeitsrechnung anbelangt, sollte man auch die Ergebnisse der Studien von Carranza & Kuzniak, (2008) ernster nehmen und die subjektivistische Interpretation von Wahrscheinlichkeit besser integrieren, damit man wäre auch für ein tiefergehendes Verständnis der Methoden der beurteilenden Statistik besser gerüstet. Das bezieht auch die enge Verbindung zwischen Wahrscheinlichkeit und Risikobetrachtungen mit ein (Borovcnik, 2016b)

Vancsó (2009, 2018) zeigt das Potential eines parallelen Zugangs zu klassischen statistischen Testmethoden und Bayes-Methoden auf, die Verschränkung beginnt bereits in der Wahrscheinlichkeitsrechnung. Migon & Gamerman (1999) gibt ein Beispiel, das auf Universitätsniveau diese Parallelität durchgehend aufbaut. Dabei wird der Begriff der bedingten Wahrscheinlichkeit zentral, Ansätze dazu gibt es in Vancsó et al. (2022). Eine profunde Analyse der Rolle der bedingten Wahrscheinlichkeit für den Unterricht findet man in Borovcnik (2012).

Gewisse Kontexte, wie die Medizin, bieten enormes Potential, Methoden der beurteilenden Statistik besser auszuloten (Borovcnik, 2019, 2020, 2022). Sie zeigen den Zweck der Modellierung mit Wahrscheinlichkeiten auf und zeigen deutlich, auf welch dünnen Füßen das Fundament der beurteilenden Statistik steht. So steht der Signifikanztest doch tatsächlich unter ständiger Kritik (Hubbard & Bayarri, 2003), sodass sich sogar die *American Statistical Association* veranlasst gesehen hat, einer Überinterpretation der Methoden entgegenzuwirken (ASA, 2016). Eine Reduktion auf die

sogenannte *Informal Inference*, welche die statistische Beurteilung allein durch *Resampling* bestehender Daten anstrebt und damit Wahrscheinlichkeit völlig überflüssig macht (und die Interpretation von Wahrscheinlichkeit auf eine primitive Häufigkeitsdeutung reduziert), bringt den letzten Rest an Logik in den Verfahren zum Verschwinden, wenn man daran denkt, beurteilende Statistik alleinig auf Resampling-Verfahren reduzieren wollte (Borovcnik, 2021a und b). Was unterrichtliche Ansätze anbelangt, so liegt in der Demonstration grundlegender Eigenschaften stochastischer Konzepte durch Applets eine gewisse Chance, Borovcnik (2018) zeigt einige Möglichkeiten auf. Interessant ist auch der Ansatz von Batanero & Borovcnik (2016), Kontexte für den Unterricht zu nutzen, in denen fundamentale Konzepte der statistischen Beurteilung eine natürliche, wenngleich auf den Kontext eingeschränkte Interpretation haben.

Ich danke Franz Pauer für seine kritischen Anmerkungen, welche die Verständlichkeit meiner Ausführungen sehr verbessert hat.

Literatur

Adam, D. (2020): Special report: The simulations driving the world's response to COVID-19. *Nature 580*(7803), 316–318. https://doi.org/10.1038/d41586-020-01003-6

Agrawal. V., Cantor, J. H., Sood, N., Whaley, C. M. (2022): *The impact of the COVID-19 pandemic and policy responses on excess mortality*. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w28930

ASA (2016): Statement on statistical significance and p-values. *The American Statistician* 70(2), 131–133. amstat.tandfonline.com/doi/full/10.1080/00031305.2016.1154108#.XKKMv9hS-Hs

Batanero, C. & Borovcnik, M. (2016): *Statistics and probability in high school*. Rotterdam: Sense Publishers.

Black, F., & Scholes, M. (1973): The pricing of options and corporate liabilities. *Journal of Political Economy 81*(3), 637–654. https://doi.org/10.1086/260062 .

Borovcnik, M. (2009): Aufgaben in der Stochastik – Chancen jenseits von Motivation. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG)*42, 20–42.

Borovcnik, M. (2012): Multiple perspectives on the concept of conditional probability. *Avances de Investigación en Didactica de la Matemática, 2*, 5–27.

Borovcnik, M. (2014): Wie artifiziell müssen Daten sein, damit sie für die Wirklichkeit relevant werden. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG)*47, 1–16.

Borovcnik, M. (2016a): Probabilistic thinking and probability literacy in the context of risk. *Educação Matemática Pesquisa, 18*(3), 1491–1516.

Borovcnik, M. (2016b): Risiko – Ein Überlebensratgeber. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG), 49*, 1–16.

Borovcnik, M. (2018): Stochastische Konzepte und Modelle mit dynamischen Applets interaktiv erforschen. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG), 51*, 1–16.

Borovcnik, M. (2019): Risk and Decision Making – Modeling and Statistics in Medicine: Fundamental Aspects. In B. Sriraman (Hrsg.): *Handbook of the Mathematics of the Arts and Sciences*. Springer Nature. doi.org/10.1007/978-3-319-70658-0_62-1

Borovcnik, M. (2020): Contexts, analogies, and tasks that expose the purpose of the key concepts of probability. *Zetetiké 28*, 1–24 –e020008

Borovcnik, M. (2021a): Informelle statistische Inferenz. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG), 53*, 1–16.

Borovcnik, M. (2021b): The meaning of probability from a foundational perspective. *Revista Sergipana de Matemática e Educação Matemática*, *6*(3), 42–76. https://doi.org/10.34179/revisem.v6i3.14904

Borovcnik, M. (2022): Modellierung und Statistik in der Medizin. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG), 54*, 1–16.

Borovcnik, M. (2023): Die Rolle der Wahrscheinlichkeit für das Verständnis beurteilender Statistik. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG), 55*, 15–30.

Borovcnik, M.; Kapadia, R. (2011): Modelling in probability and statistics – Key ideas and innovative examples. In: J. Maaß; J. O'Donoghue (Hrsg.): *Real-world problems for secondary school students – Case studies*. Rotterdam: Sense Publishers, 1–44

Borovcnik, M., Ossimitz, G. (1987): *Materialien zur Beschreibenden Statistik und Explorativen Datenanalyse* (mit Beiträgen von H. Jungwirth und B. Kröpfl). Wien: Hölder-Pichler-Tempsky.

Borovcnik, M. & Terán, T. (2023): the project method as a strategy to teach and learn applied statistics. *Proceeedings 20th International Conference on Statistical Sciences*, Vol. 37, ix–xvi. Pak Institute of Statistical Training and Research.

Borovcnik, M., & Vancsó, Ö. (2023): Die Entwicklung stochastischen Denkens. In: G. Ambrus & J. Sjuts & É. Vásárhelyi (Hrsg.): *Mathematik und mathematisches Denken – Ansprüche und Anforderungen vor, in und nach der Schule*. Münster: WTM-Verlag, 345–384. doi.org/10.37626/GA9783959872720.0.18

Cardano, G. (1560): Liber de ludo aleae. In: *Opera Omnia*, Bd. 1. Veröffentlicht 1663.

Carranza, P., & Kuzniak, A. (2008). Duality of probability and statistics teaching in French education. In C. Batanero, et al. (Hrsg.), *Proceedings of the ICMI Study 18 and 2008 IASE Round Table.* ICMI/IASE.

Joffe, R. (2021): COVID-19: Rethinking the lockdown groupthink. *Frontiers in Public Health 9*. https://doi.org/10.3389/fpubh.2021.625778

Gelman, A. (2016): Brexit polling: What went wrong? *Statistical Modeling, Causal Inference, and Social Science*. https://statmodeling.stat.columbia.edu/2016/06/24/brexit-polling-what-went-wrong/

Gelman, A., & Azari, J. (2017): 19 Things we learned from the 2016 election. *Statistics and Public Policy 4*(1), 1–10. https://doi.org/10.1080/2330443X.2017.1356775

Heisenberg, W. (1969): *Der Teil und das Ganze. Gespräche im Umkreis der Atomphysik*. München: Piper.

Hubbard, R., & Bayarri, M. J. (2003). Confusion over measures of evidence (p) versus errors (α) in classical statistical testing. *The American Statistician 57*(3), 171–182. doi.org/10.1198/0003130031856

Kahneman, D. & Tversky, A. (1979): Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–292.

Laplace, P. S. (1812): Essai philosophique sur les probabilités. *Journal de l'École Polytechnique VII/VIII*, 140–172. Englische Übersetzung 1951. New York: Dover.

Lecoutre, M. P. (1992): Cognitive models and problem spaces in „purely random" situations. *Educational Studies in Mathematics 23*(6), 557–568.

Mader, S., Rüttenauer, T. (2022): The effects of non-pharmaceutical interventions on covid-19 mortality: A generalized synthetic control approach across 169 countries. *Frontiers in Public Health 10*. https://doi.org/10.3389/fpubh.2022.820642

Migon, H. S. & Gamerman, D. (1999). *Statistical inference: An integrated approach*. London: Arnold.

Nordmeier, G. (1989): ‚Erstfrühling' und 'Aprilwetter' – Projekte in der explorativen Datenanalyse. *Stochastik in der Schule 9*(3), 21–42.

Vancsó, Ö. (2009): Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal of Mathematics Education 4*(3), 181–212. doi.org/10.29333/iejme/242

Neuwirth, E. (o.D.): *Bildung und Statistik. Statistisches Wissen (auch) zur Bildungspolitik*. https://blogs.neuwirth.priv.at/bildungundstatistik/bildungsstandards-und-pisa-aufgaben/

Statistik Austria (o.D.): Öffentliche Finanzen. *Volkswirtschaft und Öffentliche Finanzen*. https://www.statistik.at/statistiken/volkswirtschaft-und-oeffentliche-finanzen/oeffentliche-finanzen

Vancsó, Ö. (2009): Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal in Mathematics Education 4*(3), 291–322. www.iejme.com/032009/main.htm

Vancsó, Ö. (2018, Juli): How visualisation using software helps understanding classical and Bayesian statistics. *„Teaching Probability in School – Understanding and Linking it to Statistics."* ICOTS 10, Kyoto. www.researchgate.net/profile/Oedoen_Vancso

Vancsó, Ö., Fejes-Tóth, P., Borovcnik, M. (2022): Conditional probability, Bayes and classical statistics – Evaluation of the planned secondary-school reform in Hungary. In: S. Peters (Hrsg.*): Proceedings of the 11th Conference on Teaching Statistics*. The Hague: ISI.

Varga, T. (1983): Statistics in the curriculum for everybody – How young children and how their teachers react. In: Grey, D. R. et al. (Hrsg.): *Proceedings of the First Intern. Conf. on Teaching Statistics*, vol. 1. Sheffield: Teaching Statistics Trust, 71–80.

Verfasser

Manfred Borovcnik
Universität Klagenfurt, Institut für Statistik, Sterneckstraße 15, 9020 Klagenfurt
manfred.borovcnik@aau.at