REIMUND VEHLING

Beschreibende Statistik mit eigenen Daten

Vernetzungen in der Stochastik

LERNGRUPPE:	Einführungsphase
IDEE:	Ein roter Faden durch die Beschreibende Statistik mit Vernetzungen zur Wahrscheinlichkeitsrechnung und Beurteilenden Statistik
ARBEITSBLATT 1/2:	Der Weg zur empirischen Standardabweichung; Simulation eines Zufallsversuchs
WEITERES MATERIAL:	GeoGebra-Dateien
VORWISSEN:	Inhalte der Beschreibenden Statistik der Sek II
ZEITBEDARF:	ca. 6 Wochen (für die gesamte Unterrichtseinheit)

„In den nächsten Stunden arbeitet Ihr als Datendetektive - wie könnte das ablaufen?“ In meiner Klasse 11 entwickelte sich schnell eine lebhafte Diskussion mit folgenden Ergebnissen: Zuerst wird eine Forschungsfrage formuliert, dann benötigen wir Daten. Dazu müssen wir einen Experimentierplan überlegen. Die Daten werden dann erfasst und wir versuchen, die Daten zu interpretieren. Dazu benutzen wir verschiedene Darstellungen und Kenngrößen. Der Anfangsimpuls „Datendetektive“, war – bis auf Hilfen bei der Nennung einiger Fachbegriffe – fast mein gesamter Beitrag während dieser Phase, die Bezeichnung also offensichtlich tragfähig und anregend.

Mein Ziel war es, dass meine Schülerinnen und Schüler gleich in der ersten Stunde den „roten Faden“ der Unterrichtseinheit erkennen und erfahren, was sie Neues lernen und hoffentlich später auch anwenden können. Ein selbst erhobener Datensatz, auf den immer wieder zurückgegriffen werden kann, steht dabei im Zentrum.

Die Schülerinnen und Schüler nannten alle wichtigen Möglichkeiten einer Darstellung sowie viele schon aus der Mittelstufe bekannte Kenngrößen: Säulen-, Balken- und Kreisdiagramm, Spannweite, Durchschnitt, kleinster und größter Wert, Median, Modalwert. Eine Schülerin brachte sogar den Begriff „Streuung“ ins Spiel, natürlich ohne schon zu wissen, wie man so etwas messen kann. Das sind die schönen Momente im Unterricht. Ich konnte an dieser Stelle sagen: „Prima, genau darauf kommen wir später zurück und wir werden es schaffen, auch eine Formel für die „Streuung“ zu finden.“

In dieser Einheit wurden die empirische Varianz und die empirische Standardabweichung sowie die Darstellung einer Häufigkeitsverteilung als Histogramm und als Boxplot eingeführt. Es wurde auch der Einfluss der Klassenbreiten auf die Form des Histogramms untersucht.

Bei der Forschungsfrage habe ich bewusst eine (didaktische) Reduktion vorgenommen: Ich habe als Forschungsobjekt die große Handspanne vorgegeben, die definiert ist als Abstand zwischen Daumen- und Zeigefingerspitze an der aufgespannten Hand. Dieses Maß wird gelegentlich als Schätzgröße verwendet. Hierbei ist 20 cm ein guter Schätzwert. Meine Schülerinnen und Schüler haben übrigens Werte zwischen 10 cm und 30 cm als Schätzungen genannt.

Die Lernenden waren sehr kreativ und fanden viele Forschungsfragen. Und wieder eine Reduktion: Wir haben uns auf die beiden folgenden Fragen beschränkt:

- Gibt es Unterschiede zwischen linker und rechter Hand?
- Gibt es Unterschiede zwischen den Geschlechtern?

Darüber hinaus haben wir viele kreative Fragen wie „Haben Klavierspieler eine größere Handspanne?“; „Gibt es Zusammenhänge zwischen der Handspanne und dem Gewicht?“ nicht untersucht.

Es muss unbedingt Klarheit darüber bestehen, wofür die erhobenen Daten dienen: Wir wollen damit Fragen beantworten, die sich nur auf unsere Klasse beziehen, kurz: Die Gesamtheit sind in diesem Fall alle Schülerrinnen und Schüler meiner 11. Klasse. Wir befinden uns in der Beschreibenden Statistik und ermitteln jeweils Vollerhebungen. Erst in der Beurteilenden Statistik ändert sich die Sichtweise. Dann geht es z. B. um die Frage, wie groß das arithmetische Mittel der Handspanne aller Schülerinnen und Schüler der 11. Klassen in Niedersachsen ist. Unser Wert würde dann die Rolle eines Stichprobenergebnisses einnehmen, mit dem wir dann auf den wahren Wert der Grundgesamtheit mit Mitteln der Wahrscheinlichkeitsrechnung schließen, z. B. mit einem Konfidenzintervall. Ich plädiere dafür, mit der beschreibenden Statistik zu beginnen, bei der die Gesamtheit den Lernenden bekannt ist. Darauf, dass man in der Beurteilenden Statistik strenggenommen eine Zufallsstichprobe benötigt und nicht einfach die Daten irgendeiner 11. Klasse nehmen darf, ist dann erst später, bei der beurteilenden Statistik, hinzuweisen.

Black & White

Die erhobenen Messdaten (Tipp: vorher die Messgenauigkeit vereinbaren, hier 1 mm) wurden in den TI-84+ eingegeben. Ich habe mich dafür entschieden, den Rechner zuerst als „Black-Box“ zu benutzen, um dann im nächs-

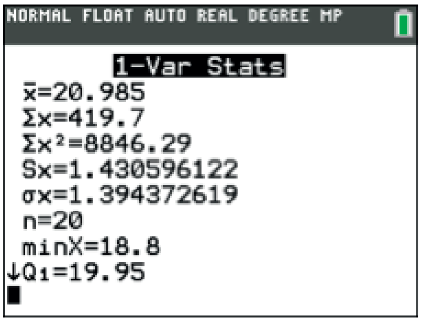


Abb. 1: Erster Teil der Black-Box

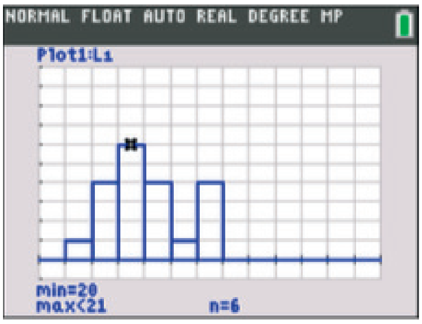


Abb. 2: Histogramm (absolut)

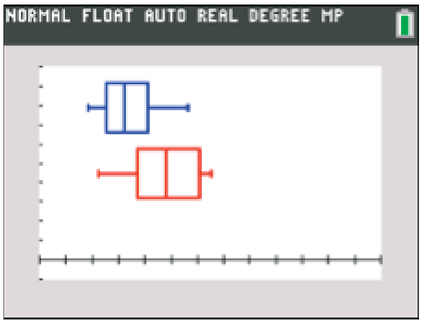


Abb. 3: Boxplots (Handspanne, w/m)

ten Schritt zur „White-Box" zu gelangen. Der Rechner liefert auf Knopfdruck die folgenden Kenngrößen **(Abb. 1)**, die nacheinander besprochen wurden: arithmetisches Mittel ($\bar{x}$), Standardabweichung der als Stichprobe aufgefassten Daten (S_x), Standardabweichung der als Grundgesamtheit aufgefassten Daten (σ_x), kleinster und größter Wert (minX, maxX), Median (Med) sowie das erste und dritte Quartil (Q_1, Q_2).

Ich habe – wie allgemein üblich – für die Standardabweichung der Grundgesamtheit den Begriff der empirischen Standardabweichung benutzt und als Bezeichnung s statt σ_x gewählt. Diese Bezeichnung findet man auch in vielen Schulbüchern.

Die unterschiedlichen Standardabweichungen und Bezeichnungen sorgen immer wieder für Verwirrung **(Kasten 1)**. Durch eine Simulation kann erfahren werden, was an der Universität dann als erwartungstreuer Schätzer hergeleitet wird (s. **Downloadmaterial**).

Meine Schülerinnen und Schüler waren zuerst Datendetektive. Sie konnten mit dem Rechner schnell Histogramme **(Abb. 2)** erstellen.

Besonders eindrucksvoll fanden sie die Darstellungen mithilfe von Box-Plots **(Abb. 3)**. Hieran konnten viele Fragen beantwortet werden, z. B.: Streuen die Handspannen der Jungen mehr als die Handspannen der Mädchen?

Hilfreich war dabei zusätzlich ein Rückgriff auf das arithmetische Mittel und die empirische Standardabweichung. Da es ihre Daten waren, zeigten die Lernenden ein wirkliches Interesse, Vor- und Nachteile der verschiedenen Darstellungen zu entdecken.

Empirische Standardabweichung

In einem Unterrichtsgang zur Beschreibenden Statistik kommt der empirischen Standardabweichung eine zentrale Rolle zu. Diese Größe kann den Weg zu den Sigma-Umgebungen und Prognose-Intervallen in der Wahrscheinlichkeitsrechnung anbahnen. Prognose-Intervalle sind wiederum ein zentrales didaktisches Hilfsmittel, um Konfidenzintervalle zu verstehen.

Die Schülerinnen und Schüler waren in der Stunde motiviert, endlich die Formel für die empirische Standardabweichung kennenzulernen, den Wert kannten sie bisher nur vom Rechner.

Bei der Erstellung des Arbeitsblattes **(Arbeitsblatt 1)** haben mich u. a. die folgenden Aspekte geleitet:

- Anknüpfen an Bekanntes (arithmetisches Mittel)
- Gefühl für die Kenngröße s
- entwickeln (Schätzwert)
- Eigenschaften der Kenngröße s sichtbar machen (z. B. die Invarianz gegenüber einer Verschiebung)
- Reduktion auf prägnante Fälle
- steigender Schwierigkeitsgrad
- Argumentationsanlässe
- Binnendifferenzierung

1 WISSENSWERT

Die empirische Varianz und die empirische Standardabweichung

Streuungsmaße liefern in der *Beschreibenden Statistik* eine Aussage über die Lage der einzelnen Daten bezüglich des Mittelwertes. Sie ergänzen den Mittelwert um eine zusätzliche Information über die zugrunde liegende Verteilung. Es gibt verschiedene Streuungsmaße, u.a. die mittlere quadratische Abweichung vom arithmetischen Mittel. Diese mittlere quadratische Abweichung der Daten $x_1, x_2, \dots, x_n$ von ihrem arithmetischen Mittel $\bar{x}$ wird **empirische Varianz** V genannt. Sie wird folgendermaßen berechnet:

$$V = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

Für die **empirische Standardabweichung s** gilt: $s = \sqrt{V}$.
Die empirische Standardabweichung s hat die gleiche Einheit wie die Daten.

In der *Beurteilenden Statistik* stellen die n Daten nicht die Grundgesamtheit dar, sondern dienen als Stichprobe, um die Kenngrößen einer Grundgesamtheit abzuschätzen. Es kann gezeigt werden, dass dann die Division durch $n - 1$ im Mittel einen bessern Schätzwert für die Standardabweichung der Grundgesamtheit liefert. Für große Werte von n ist der Unterschied aber vernachlässigbar.

Die Schätzungen für die Fälle A bis D bereiteten den Lernenden keine Probleme. Erst ab Fall E wurde es schwieriger, einen Schätzwert für die Streuung anzugeben. Aber genau das wollte ich erreichen. Interessant waren z. B. die Argumentationen für den Fall E. Die Klasse war sich zum Schluss einig, dass der Wert näher an 1 als an 3 liegen muss. Der Wert 1,5 bekam die größte Zustimmung. Es war deutlich zu spüren, dass sie endlich eine Berechnungsmöglichkeit für ein Streuungsmaß haben wollten. Die Formel sollte für A den Wert 0, für B, C und D jeweils den Wert 2 und für E einen Wert in der Nähe von 1,5 haben. Für die anderen Fälle gab es keine einheitliche Meinung.

Viele Gruppen haben den bekannten „Fehler“ gemacht und nur den Mittelwert der Differenzen zu $\tilde{x}$ betrachtet. Dass hier immer der Wert null herauskommt, fanden sie interessant. Darauf sind wir aber erst in der nächsten Stunde eingegangen und haben eine allgemeine Begründung geliefert. Ein schönes Beispiel für das Spiralprinzip. Ein Begriff – hier das arithmetische Mittel – wird im Laufe der Schulzeit mehrmals behandelt (ab Jg. 5) und auf eine höhere Stufe gehoben, ohne Vorheriges zu verwerfen.

Man kann nicht sicher sein, dass eine Gruppe auch die Abweichungen quadriert und dann hiervon den Mittelwert als mittlere quadratische Abweichung identifiziert. In dieser Stunde ist es aber in zwei Gruppen passiert. Die Schülerinnen und Schüler haben also unbewusst die Varianz bestimmt. Schnell wurde aber anhand der Fälle A–E klar, dass noch Nachbesserungen nötig waren. An dieser Stelle hilft eigentlich immer ein Hinweis auf die Einheiten, um (endlich) die Wurzel in der Formel zu erhalten. Die Schülerinnen und Schüler waren am Ende der Stunde sichtlich stolz, es geschafft zu haben. Sie haben eine Formel entdeckt, die auch wirklich für die empirische Standardabweichung benutzt wird.

Keine Gruppe hatte die Idee, mit Beträgen zu arbeiten. Dies ist aber auch eine sinnvolle Möglichkeit und darf nicht als falsch bezeichnet werden, denn es führt z. B. beim Median zum gängigen Streumaß MAD (mittlere absolute Abweichung vom Median):
$\text{MAD} = \frac{1}{n} \cdot \sum_{i=1}^{n} | x_i - \text{Median} |$.

Die mittlere quadratische Abweichung führt mit den Daten 4, 4, 4 und 8 (Fall E) auf die Funktion f mit $f(x) = 3 \cdot (4 - x)^2 + (8 - x)^2$. Die zugehörige Parabel hat den Scheitelpunkt $S(5|12)$. Es stimmt nicht zufällig die x-Koordinate des Scheitelpunktes mit $\tilde{x}$ überein. Das gilt immer. Hiermit hatten wir an einem Beispiel die Minimaleigenschaft der quadratischen Abweichungen von $\tilde{x}$ visualisiert (siehe Downloadmaterial) Mit dem GTR konnte diese schöne Eigenschaft auch für die anderen Fälle gezeigt werden. Auch habe ich erwähnt, dass die Arbeit mit Quadraten sich oft einfacher gestaltet als die Arbeit mit Beträgen, was sie sofort „eingesehen“ haben. Ich kenne zumindest keinen Lernenden, der gerne mit Beträgen rechnet.

Eigentlich wollte ich nicht auf das physikalische Pendent des Trägheitsmomentes zu sprechen kommen. Aber ein leistungsstarker Schüler brachte es auf den Punkt: „Damit werden Daten, die weiter außerhalb liegen, stärker berücksichtigt.“ Genau dies ist übrigens auch der Grund, weshalb das MAD-Streumaß seine Berechtigung hat, denn mit der Standardabweichung liefert man sich durch Messfehler bedingten Ausreißern aus.

s-Umgebungen vom Mittelwert

Nun konnten wir mit der selbst entdeckten Formel für die empirische Standardabweichung die Werte nachweisen, die der Rechner bei unseren Handspannen-Daten ermittelt hat. Aus der Black-Box ist eine White-Box entstanden. Es ist sinnvoll, sich bei der Herleitung auf den Kern zu reduzieren und zuerst mit einfachen Daten zu arbeiten.

Auch wenn die Schülerinnen und Schüler den Nutzen der empirischen Standardabweichung beim Vergleich verschiedener Datensätze erfahren hatten, gab es immer wieder die Frage, wozu man diese Kenngröße noch gebrauchen kann.

Genau an dieser Stelle können die σ-Umgebungen aus der Wahrscheinlichkeitsrechnung vorbereitet werden, indem nun die $k \cdot s$-Umgebungen von $\tilde{x}$ betrachtet werden **(Kasten 2)**. Hierbei lag der Fokus darauf, die sogenannten s-Regeln zu verstehen. Deshalb kam eine GeoGebra-Datei zum Einsatz (siehe Downloadmaterial), um nicht die einzelnen Berechnungen selbstständig durchführen zu müssen. **Abb. 4** zeigt die Ausgabe für die $1s$-Umgebung von $\tilde{x}$. Zusätzlich können in dieser Datei die Anzahl und der Anteil der Daten innerhalb dieser Umgebung sowie ein Boxplot und ein Histogramm mit variabler Klassenbreite ausgegeben werden. Da in drei weiteren Klassen Handspannenwerte erhoben wurden, hatten wir insgesamt 102 Messwerte zur Verfügung. Diese Vergrößerung des Datenpools ist wichtig, um die s-Umgebungen prägnant zu veranschaulichen.

Die s-Regeln konnten für diesen gesamten Datensatz sowie für die einzelnen Klassen gut empirisch überprüft werden. So ergaben z. B. die gesamten 102 Daten für die 1s-Umgebung einen Anteil von ca. 67 % und für die $2s$-Umgebung einen Anteil von ca. 96 %.

An dieser Stelle sollte man sich aber nicht einfach damit zufrieden geben,

2 WISSENSWERT

s-Umgebungen vom Mittelwert - Faustregeln

Falls eine Häufigkeitsverteilung ungefähr glockenförmig (also eingipflig und symmetrisch) ist, dann gilt näherungsweise für die beobachteten Daten:

- ca. 68 % fallen in die 1s-Umgebung des Mittelwertes ($[\tilde{x} - s; \tilde{x} - s]$),
- ca. 95 % fallen in die 2s-Umgebung des Mittelwertes ($[\tilde{x} - 2s; \tilde{x} - 2s]$),
- fast alle Daten fallen in die 3s-Umgebung des Mittelwertes, d. h. ab 3s fängt die Einsamkeit an.

Damit kann die empirische Standardabweichung nicht nur berechnet, sondern auch mit „bloßem Auge“ abgeschätzt werden.

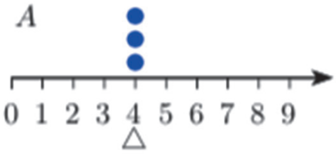
DER WEG ZUR EMPIRISCHEN STANDARDABWEICHUNG

Die Suche nach einer Berechnung für die Streuung um den Mittelwert

In den 7 Abbildungen sind Daten als Punkte dargestellt. Mit Δ wurde jeweils der Mittelwert markiert.
***Weisen** Sie nach, dass die Daten in Abbildung G tatsächlich den Mittelwert 5 haben, also $\bar{x} = 5$ gilt.*
***Geben** Sie jeweils einen Schätzwert für die Streuung um den Mittelwert **an**.*
Stellen Sie sich vor, Sie werden beauftragt, eine Berechnung für die Streuung der Daten um den Mittelwert zu entwickeln.

Seien Sie kreativ und entwickeln Sie mindestens eine Formel, mit der man die Streuung berechnen könnte.

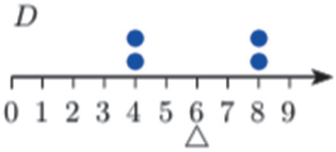
A



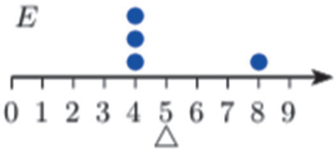
B

C

D

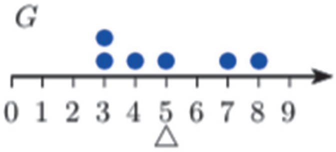


E



F

G



SIMULATION EINES ZUFALLSVERSUCHS

Wo landen ca. 95% aller Stichproben?

Es soll 36-mal eine Laplace-Münze geworfen werden. Wir gehen also davon aus, dass beim Werfen der Münze beide Seiten die gleiche Wahrscheinlichkeit besitzen. Wir erwarten also theoretisch, dass 18-mal Kopf und 18-mal Zahl erscheint. Allerdings es gibt bei einem realen Experiment immer Schwankungen. Die Variabilität des Zufalls kann man nicht ausschalten! Es gibt aber Muster, die uns weiterhelfen. Dafür muss ein Zufallsexperiment sehr oft wiederholt werden. Der Zufall lässt sich nicht so einfach in die Karten schauen.

1. Nun zu der spannenden Frage, die wir heute untersuchen werden:
Wenn wir den Versuch sehr oft durchführen, wie sieht dann wohl die Verteilung für *Anzahl Kopf* aus?

2. Es gibt immer eine Unsicherheit. Statistiker auf der ganzen Welt akzeptieren in vielen Fällen eine Unsicherheit von 5 %. Das werden wir auch machen. Sie formulieren die Frage so:
Wie sieht die Umgebung um den Erwartungswert 18 aus, in der ca. 95 % aller Stichprobenergebnisse (*Anzahl Kopf*) liegen?

   Das Problem werden wir nun mithilfe einer Simulation angehen. Unser Rechner kann würfeln! Außerdem kann er schnell verschiedene Anzahlen ermitteln.

Was will ich?	**Wie geht das mit dem TI-84+?**	**Beispiel**

Münzwurf	`MATH │ PROB │ 5:randInt`
	`lower:`
	`upper:`
	`n:`
		Münze mit Kopf/Zahl:
		`lower: 0`
		`upper: 1`
		`n: 5`
		Hiermit wird das 5-malige Werfen einer Münze simuliert (Kopf: 1; Zahl: 0).
		Beispiel (so sieht es auf den Display aus):
		`randInt(0,1,5)`
		mögliche Ausgabe: 1 0 1 1 1

summieren	`List │ MATH │ 5: sum`	`sum(randInt(0,1,36)` mögliche Ausgabe: 20 Ergebnis: 20-mal Kopf Einfach wieder Enter drücken, dann wird noch einmal 36-mal gewürfelt und die Anzahl Kopf ausgegeben.

dass sich die empirische s-Regel hier als gut brauchbar erweist. Im Unterricht zeigt sich hier ein Problem, da in einem Grundkurs oft nur die Binomialverteilung betrachtet wird. So fällt die Sigma-Regel vom Himmel, da die Normalverteilung und der zentrale Grenzwertsatz nicht behandelt werden. Es gibt in meinen Grundkursen immer wieder Schülerinnen und Schüler, die genau an dieser Stelle wissen wollen, weshalb z. B. die Faktoren 2 und 1,96 bei den Sigma-Regeln auftauchen. Die Fokussierung auf nur eine Verteilung halte ich für bedenklich. Nicht einmal die wichtige hypergeometrische Verteilung wird verbindlich vorgeschrieben. Dabei sind viele Aufgabenstellungen aufgrund einer zu geringen Grundgesamtheit gar nicht mit der Binomialverteilung modellierbar.

In der Beschreibenden Statistik können relativ einfach diese s-Regeln empirisch bei einigen Häufigkeitsverteilungen überprüft werden. Wichtig ist, dass die Schülerinnen und Schüler hierbei erkennen, wieso diese Regel bei einigen Häufigkeitsverteilungen gute und bei anderen Verteilungen nicht so gute Ergebnisse liefert. Ausgehend von verschiedenen Beispielen können Schülerinnen und Schüler selbstständig darauf kommen, dass eingipflige und möglichst symmetrische Häufigkeitsverteilungen die empirische s-Regel oft (nicht immer!) gut erfüllen, während mehrgipflige oder schiefsymmetrische Verteilungen eher schlechte Ergebnisse erzielen.

Eigentlich benutze ich an dieser Stelle gerne die Temperaturdaten vom Central-Park, die seit 1869 monatlich veröffentlicht werden. Hier ergeben sich mehrgipflige Häufigkeitsverteilungen, bei denen die s-Regeln nicht gelten. Aber diesmal hatte ich mich aus Zeitgründen nur für „konstruierte" Messwerte entschieden: Die 16 natürlichen Zahlen von 11 bis 30 sollten Messwerte darstellen. Es liegt also eine Gleichverteilung vor. Auch ohne Rechnung konnte schon vermutetet werden, dass die s-Regeln hier nicht gelten. Mit $\tilde{x} = 20{,}5$ und $s \approx 5{,}77$ liegen nur 60 % der Daten in der 1s-Umgebung, aber 100 % in der 2s-Umgebung.

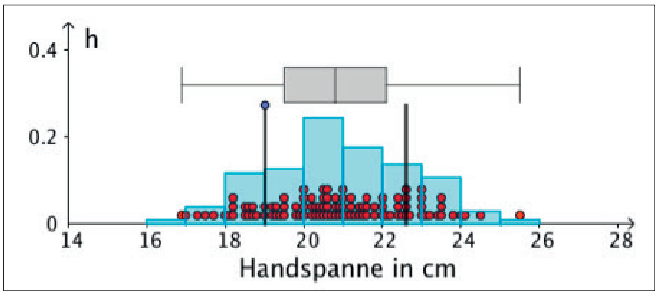


Abb. 4: Boxplot, Histogramm, PunktePlot. Die 1s-Umgebung von x ist durch die beiden grauen Striche bei 19 und 23 begrenzt

Das Tor zur Wahrscheinlichkeitsrechnung

Zum Abschluss der Unterrichtseinheit öffne ich mit der folgenden Aufgabe (s. **Arbeitsblatt 2**) die Tür zur Wahrscheinlichkeitsrechnung:

→ *Eine Laplace-Münze wird 36-mal geworfen. Wie sieht wohl die Häufigkeitsverteilung für Anzahl Kopf aus, wenn wir den Versuch oft durchführen?*
Wie groß ist die Umgebung um den Mittelwert, in der ca. 95 % aller Stichproben der Länge 36 liegen?

Die Aufgabenstellung sollte zum einen den Blick in die Zukunft ermöglichen und zum anderen „sanft" auf Stichprobenverteilungen und die Vernetzung von Beschreibender Statistik und Wahrscheinlichkeitsrechnung vorbereiten.

Auch hier ist es wichtig, vor dem Rechnen bzw. Simulieren erst einmal Hypothesen einzufordern. An dieser Stelle können meine Schülerinnen und Schüler nur mithilfe einer Simulation die Aufgabe lösen, da die Binomialverteilung natürlich noch nicht bekannt ist. Die Simulation erfolgte mit dem TI-84+ mithilfe des Befehls sum(randInt(0,1,36), der jeweils 20-mal durchgeführt wird. Alle Lernenden haben hiermit 20 Stichproben der Länge 36 erzeugt. Dieses Klassenexperiment ist einer einzelnen Computersimulation, die mit GeoGebra auf Knopfdruck realisiert werden kann, vorzuziehen. Eine GeoGebra-Datei wird – wie bei fast allen Klassenexperimenten – für die Aufnahme der einzelnen Stichprobenergebnisse, die grafische Veranschaulichung und die Berechnungen benutzt (**Abb. 5** und Downloadmaterial). Meinen Hinweis, dass wir in der Wahrscheinlichkeitsrechnung hierfür ein mathematisches Modell kennenlernen werden, mit dem man statt unserer Stichprobenverteilung eine Wahrscheinlichkeitsverteilung erhält, fanden sie interessant. Aber noch interessanter fanden sie es, als ich ihnen mitteilte, dass die (theoretische) Standardabweichung in diesem Modell exakt den Wert 3 hat – unser Wert lag nur knapp daneben. Einige fanden es spannend, dass man in der Mathematik so etwas „Zufälliges" auch „berechnen" kann. Dass der Mittelwert $\tilde{x}$ bei $n \cdot p = 18$ liegt, wurde natürlich sofort vermutet: „Was denn sonst?"

Die Problemstellung ist m. E. gut geeignet, am Ende der Unterrichtseinheit einen Ausblick auf die Wahrscheinlichkeitsrechnung zu geben. Wichtig finde ich aber auch zu hinterfragen, wofür man eigentlich die Antworten gebrauchen kann. Ein kurzer Lehrervortrag zum Begriff random walk, sowie zur Modellierung von Teilchenbewegungen im Raum schlossen sich deshalb an.

Rückblick

Wir haben wenig mit dem eingeführten Schulbuch gearbeitet. Das kann durchaus kritisch gesehen werden. Aber gerade bei der Beschreibenden Statistik ist es möglich, mit eigenen Daten den Unterricht für Schülerinnen und Schülern spannender zu gestalten. Die Frage „Wofür brauche ich das?" habe ich nicht gehört. Zu Beginn der Unterrichtseinheit habe ich übrigens die Frage gestellt, ob wir mit den Aufgaben aus dem Buch, Daten aus Erhebungen oder eigenen Daten arbeiten sollen. Die Antwort fiel eindeutig zugunsten der Arbeit mit eigenen Daten aus. Es muss nicht immer eine aufwendige Datenerhebung sein. Auch das Messen der Handspanne führt zu den zentralen Punkten der Beschreibenden Statistik. Diese Daten haben uns während der gesamten Unterrichtseinheit begleitet. Sie bildeten den berühmten „roten Faden".