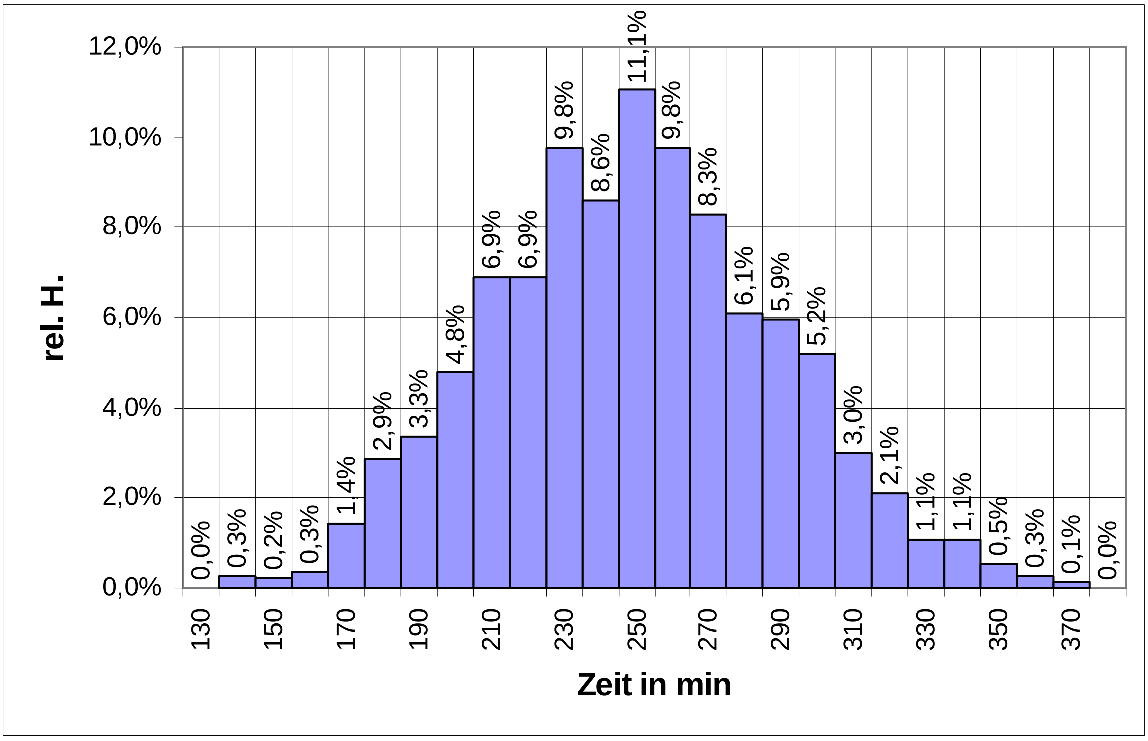
Beschreibende Statistik: GeoGebras Kalkulationstabellen als Werkzeug und als Lernumgebung nutzen

Wolfgang Riemer, Günter Seebach



Glockenförmige Verteilung der Zeiten von 1464 „Köln-Marathon" Läufern mit Mittelwert μ^=249,6 und Standardabweichung σ^=39,6 (in Minuten).

Im folgenden Beitrag

- wird GeoGebra als Werkzeug zur Aufbereitung und Visualisierung realer Daten genutzt,
- wird durch statistische Untersuchung von Messfehlern Geometrie mit beschreibender Statistik verknüpft, um im Unterricht „zwei Fliegen mit einer Klappe zu schlagen",
- werden Simulationen genutzt, um theoretische Zusammenhänge („geniale" Eigenschaften der Standardabeichung) schon in der Sekundarstufe 1 aufzudecken,
- erkennt man: Die Software entlastet ***als Werkzeug*** nicht nur von Routine sondern ermöglicht ***als Lernumgebung*** auch das Entdecken tief liegender Zusammenhänge.

Ziele

Durch Untersuchen von Daten gewinnt man Einsichten. Dazu lernt man in der beschreibenden Statistik, Daten graphisch aufzubereiten, durch Kennwerte zu beschreiben und im Hinblick auf vorab formulierte Fragestellungen zu interpretieren. Die nötigen Arbeitsschritte bei der Datenaufbereitung bringen, wenn man sie einmal verstanden hat, keinen Erkenntnisgewinn. Sie lenken vom viel wichtigeren Interpretieren ab. Deswegen erledigt man diese Routine heute mit Tabellenkalkulationsprogrammen. Geogebra bietet eine Alternative zu klassischen Werkzeugen wie Excel, OpenOffice oder graphischen Taschenrechnern. Den Leser davon zu überzeugen und ihm den Einstig in GeoGebra zu erleichtern -

ist das **erste** Ziel dieses Beitrages. **Zweitens** wollen Beispiele bewusst machen, wie fruchtbar es sein kann, beschreibende Statistik auch einmal in den Dienst der Geometrie zu stellen. Man untersucht Messungenauigkeiten (Dreiecksflächen, Bestimmung von Pi) statistisch und gewinnt nicht nur ein Gefühl für sorgfältiges Arbeiten, sondern auch für den prinzipiellen Unterschied zwischen exakter Mathematik (Modellwelt, Theorie) und vermessener Wirklichkeit. **Drittens** sollen Simulationen genutzt werden, um „geniale" Eigenschaften der Standardabweichung zu beleuchten - und um so schon im Rahmen beschreibender Statistik der Sekundarstufe I Freundschaft mit der Normalverteilung zu schließen. Die Kalkulationstabellen und Kopiervorlagen sind im Download erhältlich.

Mittelwert, Median, Balkendiagramm und Boxplot

Viele Lehrpläne fordern die Behandlung *sowohl* des Mittelwerts (arithmetisches Mittel) *als auch* des Medians zur Beschreibung von Datensätzen. Dafür gibt es mehrere Gründe:

- Zur Bestimmung des Medians muss man nur ordnen und abzählen, kaum rechnen, das geht schon in der Primarstufe.
- Median und Mittelwert beantworten *völlig unterschiedliche* Fragen.
- Den Median benötigt man, um Boxplots zu verstehen. Boxplots fassen dabei die in den Daten steckenden Informationen noch übersichtlicher zusammen als Balkendiagramme.
- Mit der Länge der Box (dem Quartilabstand) besitzt man ein einfach zu interpretierendes Streumaß!
- Die Standardabweichung, die die Streuung um den Mittelwert misst, ist dagegen wegen der auftretenden Wurzel nicht nur schwerer zu berechnen. Ihre inhaltliche Deutung (68%-Regel, die nur bei glockenförmigen Verteilungen gilt), ist schwieriger und letztlich nur im Zusammenhang mit der Normalverteilung und dem zentralen Grenzwertsatz befriedigend möglich.

GeoGebra als Werkzeug: Dein Smartphone im Fokus von Mittelwert und Median

Auf der Suche nach Daten, deren Untersuchung für Kinder höchst spannend ist und bei denen Mittelwert und Median sich sehr deutlich voneinander unterscheiden, wird man beim Smartphones fündig.

Erhalten ▽	Größe
23.04.2012 11:33	5 KB
23.04.2012 11:30	3 KB
23.04.2012 09:36	4 KB
21.04.2012 14:50	36 KB
20.04.2012 11:49	8 KB
20.04.2012 10:06	19 KB
20.04.2012 08:19	115 KB
20.04.2012 00:42	9 KB
20.04.2012 00:26	335 KB
19.04.2012 23:50	3 KB
19.04.2012 20:55	84 KB

Die Größen von Emails sind nämlich höchst unsymmetrisch verteilt: Die meisten Mails sind klein, es gibt aber auch einige sehr große. Die untere Antenne des Boxplots der die Verteilung der Größe visualisiert, endet nur wenig oberhalb von 0 kB. Der anderen Antenne, die beim Maximum endet, sind nach oben hin keine Grenzen gesetzt. Analoges gilt für die Zeitabstände zwischen eingehenden Mitteilungen und Telefongesprächen. Was man prinzipiell erwarten kann wenn Schüler die Größe der Emails in ihrem Posteingang untersuchen, zeigt Abb. 1:

- Der Median der Email-Größe hat hier den Wert 38 kB. Die Hälfte aller Emails ist also weniger als 38 kB groß, die andere Hälfte ist größer als 38 kB.
- Die mittlere Größe beträgt dagegen stolze 271 kB, das bedeutet: wenn man die gesamte eingehende Information auf die Meldungen gleichmäßig verteilen müsste, hätte jede Mail die Größe 271 kB.

Wenn man weiß, wie viele Mails man bekommt und deren mittlere Größe kennt, kann man die Verbindungskosten abschätzen. Die Kenntnis des Medians nützt bei einer solchen Abschätzung nichts.

In Kasten 1 wird erläutert, wie man größere Datenmengen auswertet und visualisiert (**Werkzeugcharakter** von GeoGebra). Diesen Kasten kann zusammen mit Abb. 1 problemlos auch Schülern als „Gebrauchsanweisung“ aushändigen. Über den Werkzeugcharakter hinaus erweist sich der aus der Geometrie geläufige Einsatz **dynamischer Arbeitsblätter** auch in der Stochastik als sehr nützlich:

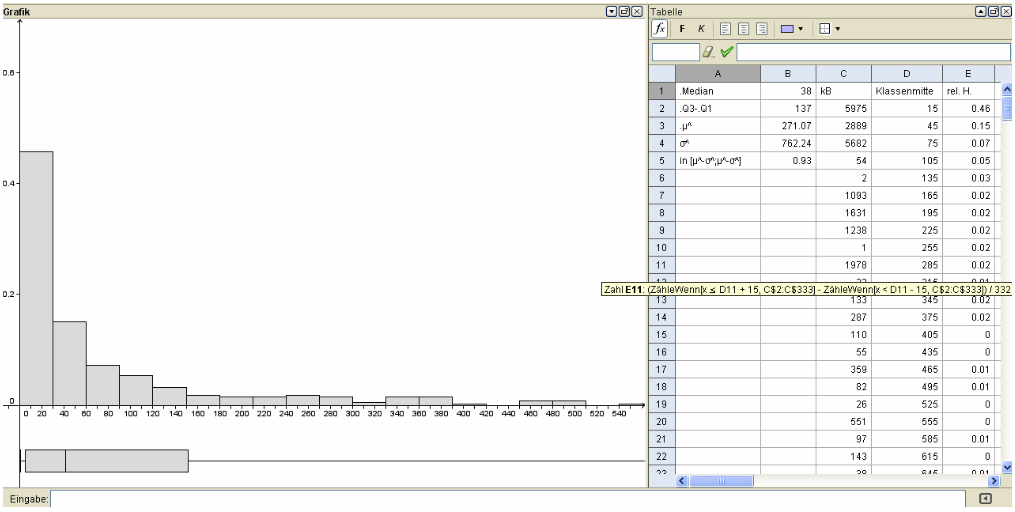


Abb. 1 Verteilung der Größe (in kB) eingehender Mails

Auswertung statistischer Daten mit GeoGebra (vgl. Abb. 1)

a) Man notiert die Daten in einer Spalte, hier die Größe von 332 Emails in Spalte C im Bereich C2:C333.

b) Man fasst diesen Bereich zu einer Liste zusammen durch L=C2:C333 in der Eingabezeile

c) In Spalte D werden die Mittelpunkte der Klassen notiert, die zur Auswertung genutzt werden sollen. Hier beträgt die Klassenbreite 30kB. Dazu kann man die ersten beiden Einträge markieren und bei gedrückter rechter Maus nach unten ziehen. Dadurch wird die Reihe der Klassenmitten nach unten fortgesetzt.

d) In Spalte E werden die Urdaten aus Spalte C ausgezählt. Die relative Häufigkeit, die auf die Klasse entfällt, erhält man z. B. in Zelle E11 über den (eingeblendeten) Befehl: (ZähleWenn(D11-15<x<=D11+15,L)/332, den man in die restlichen Zeilen von Spalte E kopiert.

d) Das Balkendiagramm wird gezeichnet durch den Befehl BalkenDiagramm[L,D2:D30]. Alternativ kann man ohne vorheriges Auszählen auch mit dem Befehl BalkenDiagramm[L,30] arbeiten. Dann werden die Klassenmitten von Geogebra automatisch festgelegt und nur die Klassenbreite wird mit dem Wert 30 vorgegeben

e) Der Boxplot entsteht durch den Befehl BoxPlot[L,-0.1,0.05], wobei der zweite Parameter -0.1 die horizontale Position der Box im Koordinatensystem und der dritte Parameter 0.05 die Breite der Box festlegen.

f) für die Kennwerte der Verteilung gibt es die Befehle Mitttelwert[L], Median[L], Q1[L] (unteres Quartil), Q3[L] (oberes Quartil) und Standardabweichung[L].

Kasten 1 Gebrauchsanweisung zur Datenauswertung mit GeoGebra

GeoGebra als Lernumgebung: ein dynamisches Arbeitsblatt zum Thema "wie sich Mittelwert, Median und Boxplots verändern, wenn..."

Mit GeoGebra ist die Entwicklung dynamischer Arbeitsblätter einfacher als in reinen Kalkulationsprogrammen. Denn man kann Punkte auf dem Zeichenblatt verschieben - und deren Koordinaten simultan im Kalkulationsblatt auswerten. Mit geeigneten Arbeitsaufträgen können Schüler Begriffe und funktionale Zusammenhänge handelnd erforschen. Durch Verbalisieren und Begründen ihrer Entdeckungen ergeben sich fruchtbare mathematische Gesprächsanlässe. Man muss als Lehrer viel weniger erklären und kann sich auf das Zusammenfassen, Strukturieren und das Sichern von Zusammenhängen konzentrieren.

Ein Beispiel ist Abb. 2 in Kombination mit Kasten 2 zu entnehmen. Durch systematisches Variieren einzelner Daten, die auf der x-Achse als Punkte vorliegen, erkennt man, wie sich das arithmetische Mittel, der Median, der Boxplot und das Balkendiagramm ändern. Man spricht über funktionale Abhängigkeiten. Die Stabilität des Medians und des Quartilabstandes gegen Ausreißer kann man kaum schöner entdecken.

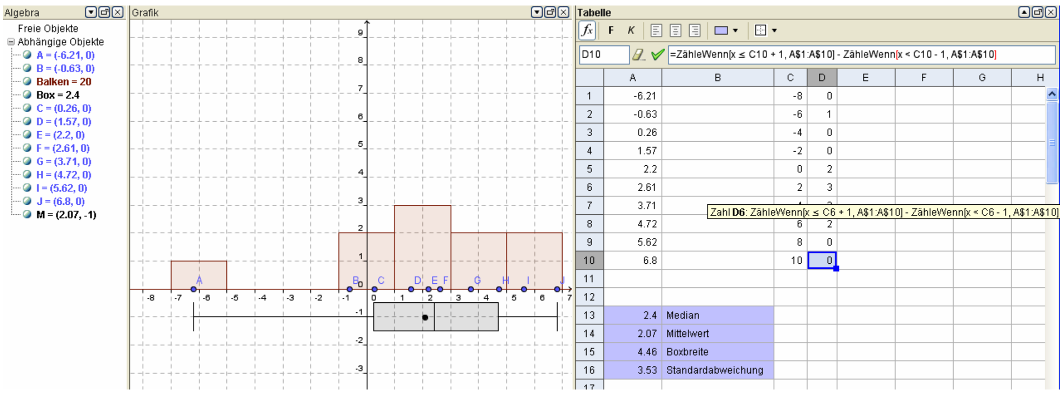


Abb. 2: Dynamisches Arbeitsblatt zum Zusammenhang zwischen Mittelwert, Median, Balkendiagramm und Boxplot. Die „technischen Details" werden im Kasten 3 erläutert.

In Abb. 2 kannst du die 10 Daten eines Datensatzes (dargestellt durch die Punkte A... I auf der Rechtsachse), die für das Taschengeld von 10 Schülern stehen könnten, durch Ziehen mit der Maus verändern. Dadurch verändern sich Mittelwert, Median. ... und natürlich auch der Boxplot und das Säulendiagramm. Schreibe auf,

a) wie sich Mittelwert, Median und Boxplot ändern, wenn man einen Datenpunkt um 2 Einheiten vergrößert,

b) wie sich dabei das Säulendiagramm verändert,

c) wie der Boxplot aus den 10 Daten hervorgeht.

d) Man sagt: Der Median ist stabil gegenüber Ausreißern (das sind besonders große/kleine Daten im Datensatz). Erläutere, was diese Aussage bedeuten könnte.

Kasten 2: Aufgabenstellung zum dynamischen Arbeitsblatt aus Abb. 2

Das dynamische Arbeitsblatt aus Abb. 2 erstellt man wie folgt:

a) Man legt 10 Punkte, A, ..., J, auf die x-Achse .

b) Mit dem Befehl x(A) erhält man die x-Koordinate von A in der Zelle A1 des Kalkulationsblattes,... , mit x(J) in Zelle A10 die x-Koordinate von J. Damit stehen die 10 über die Maus veränderbare Daten in Spalte A.

c) Das Säulendiagramm, den Boxplot und die zugehörigen Kennwerte erhält man wie in Kasten 1 erläutert.

d) Mit dem Eintrag M=(A14,-1) wird der in der Zelle A14 berechnete Mittelwert als Punkt im Boxplot visualisiert.

Kasten 3: technische Details Aufgabenstellung zum dynamischen Arbeitsblatt aus Abb. 2

Geometrie statistisch gesehen: Glockenförmigere Verteilungen, Standardabweichung

Wenn man die „Idee des Messens" als eine der fundamentalen Ideen des Mathematikunterrichts ernst nimmt, dann kommt man an der Untersuchung von Messfehlern und an der Standardabweichung als Streuungsmaß um den Mittelwert nicht vorbei. In der Tat fallen bei jeder Aufgabe, bei der im Unterricht oder in der Klassenarbeit gemessen oder mit Messwerten gerechnet wird, massenweise Daten an, deren Untersuchung auf Interesse stößt, da es doch die eigenen Daten sind. Es sind Daten, die man namentlich zuordnen - und die man untereinander vergleichen kann. Wenn man Statistik und Geometrie streng voneinander trennt, verschenkt man in der Verbindung beider Bereiche schlummerndes Motivationspotential, wie die folgenden Beispiele aus der Mittelstufe zeigen.
Aber auch in der Oberstufe gibt es Ansatzpunkte für statistische Untersuchung von Messfehlern: Wenn man z. B. vor einem analytischen Zugriff die Steigung an einen Funktionsgraphen ($f(x)= x^3$ oder $f(x)=2^x$) nach Augenmaß durch Anlegen eines Geodreiecks messen lässt – und die von der gesamten Lerngruppe ermittelten Werte statistisch auswertet – oder wenn man im Rahmen der Integralrechnung „krumme Flächen" ausmessen lässt. Die Frage nach dem genauen (analytisch zu bestimmenden) Wert ist ungleich spannender, wenn man zuvor an die Grenzen des Messens gestoßen ist.

Formel der Standardabweichung
Wenn μ^ den Mittelwert der Zahlen $x_1, x_2, ..., x_{10}$ bezeichnet, dann berechnet sich die Standardabweichung nach der Formel $\sigma\wedge = \sqrt{\frac{1}{10}((x_1-\mu\wedge)^2+(x_2-\mu\wedge)^2+...+(x_{10}-\mu\wedge)^2)}$

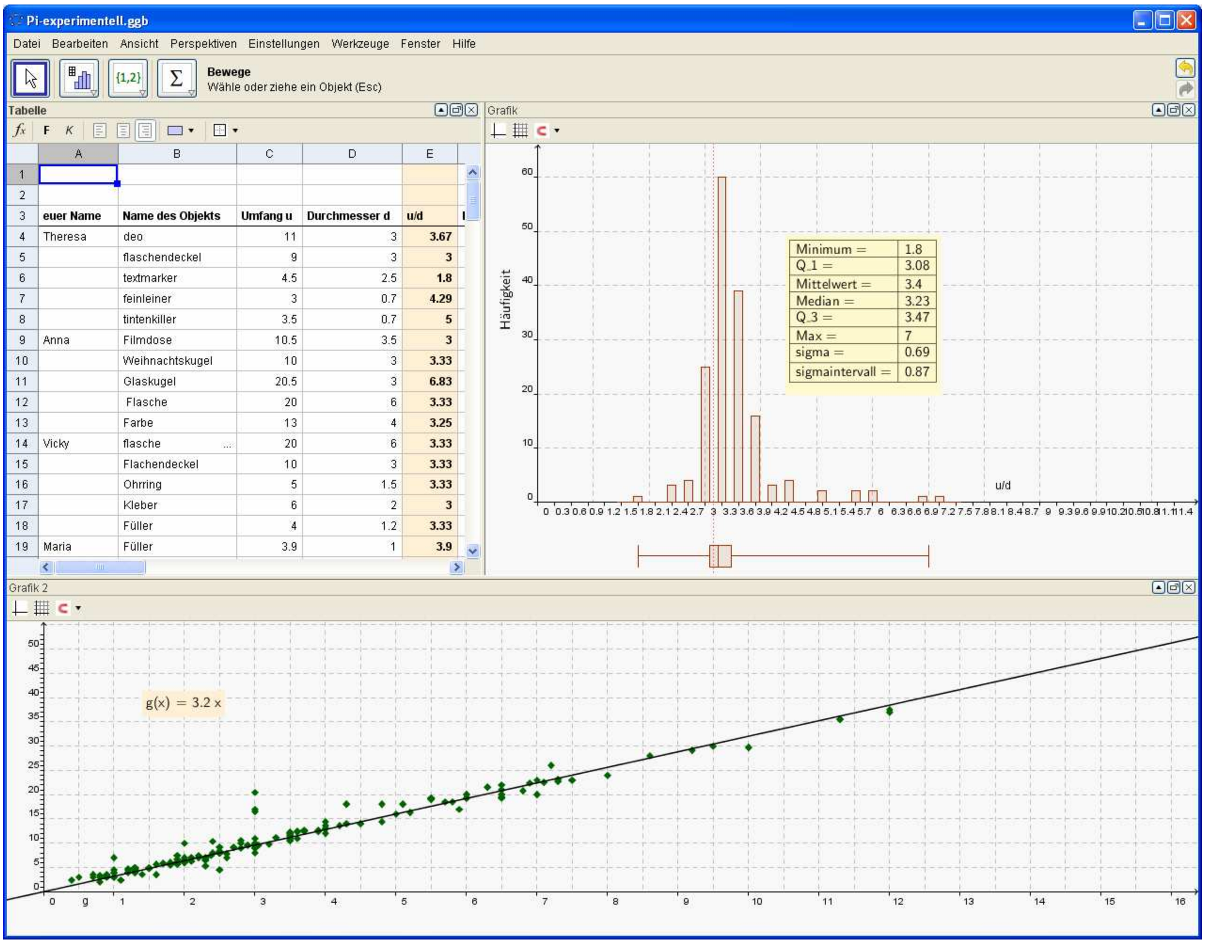
In der Formel der Standardabweichung, die man in aller Regel vorgibt oder als Black Box am Rechner erforschen lässt, erkennt man zwar sofort, dass man einen umso größeren Wert erhält, je stärker die Daten um den Mittelwert μ^ schwanken. Ihre Bedeutung ergibt sich aber erst aus der

Sigmaregel:
Wenn eine Häufigkeitsverteilung Glockenform hat (die Daten „normalverteilt" sind), dann liegen ca. 68% der Daten im Standardabweichungs-Intervall [μ^-σ^; μ^-σ^] .

So liegen in der Eingangsgrafik 73% der Laufzeiten im Sigmaintervall um den Mittelwert. Wenn Schüler *ihre eigenen* Daten auf Glockenform untersuchen – und kritisch prüfen, wie gut die „Sigmaregel" bei ihren Daten gilt, sind sie interessiert dabei. An zwei Beispielen soll erläutert werden, mit welchen Ergebnissen man zu rechnen hat.

Beispiel 1: Experimentelle Bestimmung von Pi

In der 7a stehen lineare Zusammenhänge auf der Tagesordnung. Jeder bringt fünf kreisrunde Gegenstände mit und bestimmt den Durchmesser d, den Umfang u und diktiert seine Daten in eine gemeinsame Tabelle. Die Paare (d,u) werden als Punkte geplottet. Man entdeckt einen linearen Zusammenhang, der hier gut durch die Ursprungs-Trendgerade u=3,202d beschrieben wird, die Verteilung der Quotienten u/d ist in Abb. 3 zu sehen. Hier liegen mit 87% mehr Daten im Sigmaintervall als die Faustregel angibt. Einen Ausschnitt aus den in der 7a gesammelten Urdaten zeigt Abb. 3



	A	B	C	D	E
3	euer Name	Name des Objekts	Umfang u	Durchmesser d	u/d
4	Theresa	deo	11	3	3.67
5		flaschendeckel	9	3	3
6		textmarker	4.5	2.5	1.8
7		feinleiner	3	0.7	4.29
8		tintenkiller	3.5	0.7	5
9	Anna	Filmdose	10.5	3.5	3
10		Weihnachtskugel	10	3	3.33
11		Glaskugel	20.5	3	6.83
12		Flasche	20	6	3.33
13		Farbe	13	4	3.25
14	Vicky	flasche	20	6	3.33
15		Flachendeckel	10	3	3.33
16		Ohrring	5	1.5	3.33
17		Kleber	6	2	3
18		Füller	4	1.2	3.33
19	Maria	Füller	3.9	1	3.9

Abb. 3 Messwerte für Pi (Ausschnitt), Verteilung des Quotienten u/d sowie linearer Zusammenhang zwischen d und u

So zeichnet man Trendgeraden in GeoGebra:

a) Man erzeugt eine Liste L von Punkten: $L=\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$

b) Man erstellt einen Schieberegler s und mit diesem eine Funktion f durch: f(x) = s*x in der Eingabezeile.

c) Nun erhält man die Trend-Ursprungsgerade durch Eingabe von: g(x) =Trend[L,f] in der Eingabezeile. Trend berechnet die Minimale-Fehlerquadrat-Funktion durch die Punkte in der Liste. Dabei könnte f(x) auch eine beliebige andere Funktion sein, die aber mindestens von einem Schieberegler abhängt.

Kasten 4:Technische Details zu Trendgeraden in Geogebra

Beispiel 2: Dreiecksflächenstatistik

In einer Klassenarbeitsaufgabe sollten die drei Höhen eines auf Blanco-Papier vorgegebenen stumpfwinkligen Dreiecks (Abb. 4) „so sauber wie möglich“ konstruiert werden. Anschließend war die Dreiecksfläche zu bestimmen. - und zwar dreimal über drei verschiedene Produkte

$$A = \frac{a \cdot h_a}{2} = \frac{b \cdot h_b}{2} = \frac{c \cdot h_c}{2}.$$

Das dahinter steckende Ziel ist einerseits Selbstkontrolle, andererseits die Entwicklung eines Gefühls für Mess- und Rundungsfehler. (Aller Theorie zum Trotz schneiden sich die Höhen auf dem Papier nur selten in einem Punkt. Die Dreiecksflächen weichen voneinander ab, sie sind selten identisch.)

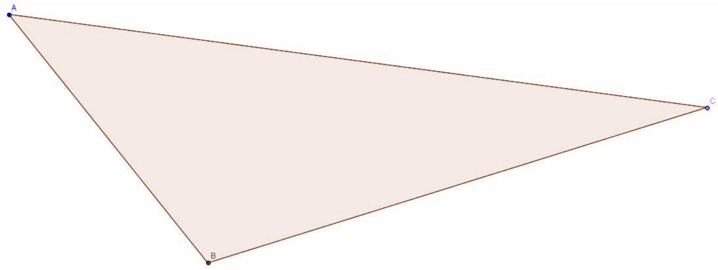
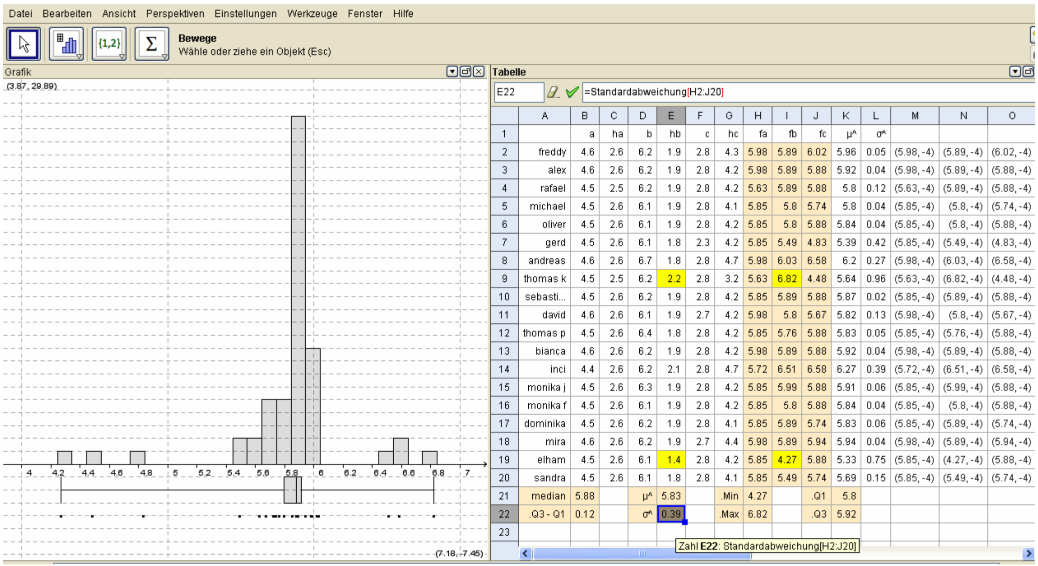


Abb.4: stumpfwinkliges Dreieck auf Blanco-Papier



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		a	ha	b	hb	c	hc	fa	fb	fc	μ^	σ^			
2	freddy	4.6	2.6	6.2	1.9	2.8	4.3	5.98	5.89	6.02	5.96	0.05	(5.98, -4)	(5.89, -4)	(6.02, -4)
3	alex	4.6	2.6	6.2	1.9	2.8	4.2	5.98	5.89	5.88	5.92	0.04	(5.98, -4)	(5.89, -4)	(5.88, -4)
4	rafael	4.5	2.5	6.2	1.9	2.8	4.2	5.63	5.89	5.88	5.8	0.12	(5.63, -4)	(5.89, -4)	(5.88, -4)
5	michael	4.5	2.6	6.1	1.9	2.8	4.1	5.85	5.8	5.74	5.8	0.04	(5.85, -4)	(5.8, -4)	(5.74, -4)
6	oliver	4.5	2.6	6.1	1.9	2.8	4.2	5.85	5.8	5.88	5.84	0.04	(5.85, -4)	(5.8, -4)	(5.88, -4)
7	gerd	4.5	2.6	6.1	1.8	2.3	4.2	5.85	5.49	4.83	5.39	0.42	(5.85, -4)	(5.49, -4)	(4.83, -4)
8	andreas	4.6	2.6	6.7	1.8	2.8	4.7	5.98	6.03	6.58	6.2	0.27	(5.98, -4)	(6.03, -4)	(6.58, -4)
9	thomas k	4.5	2.5	6.2	2.2	2.8	3.2	5.63	6.82	4.48	5.64	0.96	(5.63, -4)	(6.82, -4)	(4.48, -4)
10	sebasti...	4.5	2.6	6.2	1.9	2.8	4.2	5.85	5.89	5.88	5.87	0.02	(5.85, -4)	(5.89, -4)	(5.88, -4)
11	david	4.6	2.6	6.1	1.9	2.7	4.2	5.98	5.8	5.67	5.82	0.13	(5.98, -4)	(5.8, -4)	(5.67, -4)
12	thomas p	4.5	2.6	6.4	1.8	2.8	4.2	5.85	5.76	5.88	5.83	0.05	(5.85, -4)	(5.76, -4)	(5.88, -4)
13	bianca	4.6	2.6	6.2	1.9	2.8	4.2	5.98	5.89	5.88	5.92	0.04	(5.98, -4)	(5.89, -4)	(5.88, -4)
14	inci	4.4	2.6	6.2	2.1	2.8	4.7	5.72	6.51	6.58	6.27	0.39	(5.72, -4)	(6.51, -4)	(6.58, -4)
15	monika j	4.5	2.6	6.3	1.9	2.8	4.2	5.85	5.99	5.88	5.91	0.06	(5.85, -4)	(5.99, -4)	(5.88, -4)
16	monika f	4.5	2.6	6.1	1.9	2.8	4.2	5.85	5.8	5.88	5.84	0.04	(5.85, -4)	(5.8, -4)	(5.88, -4)
17	dominika	4.5	2.6	6.2	1.9	2.8	4.1	5.85	5.89	5.74	5.83	0.06	(5.85, -4)	(5.89, -4)	(5.74, -4)
18	mira	4.6	2.6	6.2	1.9	2.7	4.4	5.98	5.89	5.94	5.94	0.04	(5.98, -4)	(5.89, -4)	(5.94, -4)
19	elham	4.5	2.6	6.1	1.4	2.8	4.2	5.85	4.27	5.88	5.33	0.75	(5.85, -4)	(4.27, -4)	(5.88, -4)
20	sandra	4.5	2.6	6.1	1.8	2.8	4.1	5.85	5.49	5.74	5.69	0.15	(5.85, -4)	(5.49, -4)	(5.74, -4)
21	median	5.88		μ^	5.83		.Min	4.27		.Q1	5.8				
22	.Q3 - Q1	0.12		σ^	0.39		.Max	6.82		.Q3	5.92				
23															

Abb. 5 : stumpfwinkliges Dreieck auf Blanco-Papier von jedem Schüler dreimal vermessen. Die Box liegt zwischen Q_1=5,8 und Q_3=5,92. Hier liegt die mittlere Hälfte aller Messwerte. Die Länge 0,12 der Box ist ein Maß für die Messungenauigkeit, die Ausreißer unberücksichtigt lässt.

Was bei den Messungen herauskam, zeigt die statistische Auswertung in Abb. 5: 19 Schüler erhielten für die Dreiecksfläche 57 Ergebnisse. Die Verteilung der Dreiecksflächen ist als Säulendiagramm, als Boxplot und als Punktdiagramm veranschaulicht. Die Messwerte schwanken zwischen 4,27 cm² (Elham hat die Höhe h_b falsch abgelesen) und 6,82 (Thomas K hat ebenfalls die Höhe h_b falsch abgelesen). Der Mittelwert aller Dreiecksflächen beträgt 5,83 cm², die Standardabweichung 0,39 cm². Im Sigmaintervall liegen mit 88% mehr als die durch die Faustregel nahe gelegten 68% aller Daten. Die durch Ablesesefehler bedingten Ausreißer vergrößern die Standardabweichung sehr, so dass dann auch bei glockenförmigen Verteilungen mehr als 68% der Daten im Sigmaintervall zu liegen kommen.

Standardabweichung mit Simulationen erforschen

Die Standardabweichung hat zwei geniale Eigenschaften:

(1) Sie ist als Streumaß additiv. Beim Summieren vieler voneinander unabhängiger Zufallsgrößen U=X+X+ ... + X (n Summanden) entsteht eine Summengröße U, deren Standardabweichung sich aus den Standardabweichungen der Summanden berechnen lässt. Es gilt $\sigma_U = \sqrt{n} \cdot \sigma_X$.
Wenn man die Standardabweichung eines Summanden kennt, dann kennt man also auch die Standardabweichung der Summe. *Eine entsprechende Aussage gilt für den Quartilabstand nicht.*

(2) Die Verteilung von U wird bei hinreichend vielen Summanden glockenförmig („Normalverteilung“), egal wie die Summanden verteilt sind. Es gilt für die Summe die 68%-Sigmaregel.

Diese Zusammenhänge kann man schon im Rahmen beschreibender Statistik in der Sekundarstufe I aufdecken, wenn man Simulationen mit GeoGebras Zufallszahlen durchführt. Hierzu zunächst zwei Beispiele:

Beispiel 3: Die Augensumme beim zweifachen Würfelwurf

Obwohl die Augenzahlen beim einfachen Würfelwurf gleichverteilt ist, ist die Augensumme beim zweifachen Würfelwurf bereits auf dem Weg zu einer glockenförmigen Verteilung.

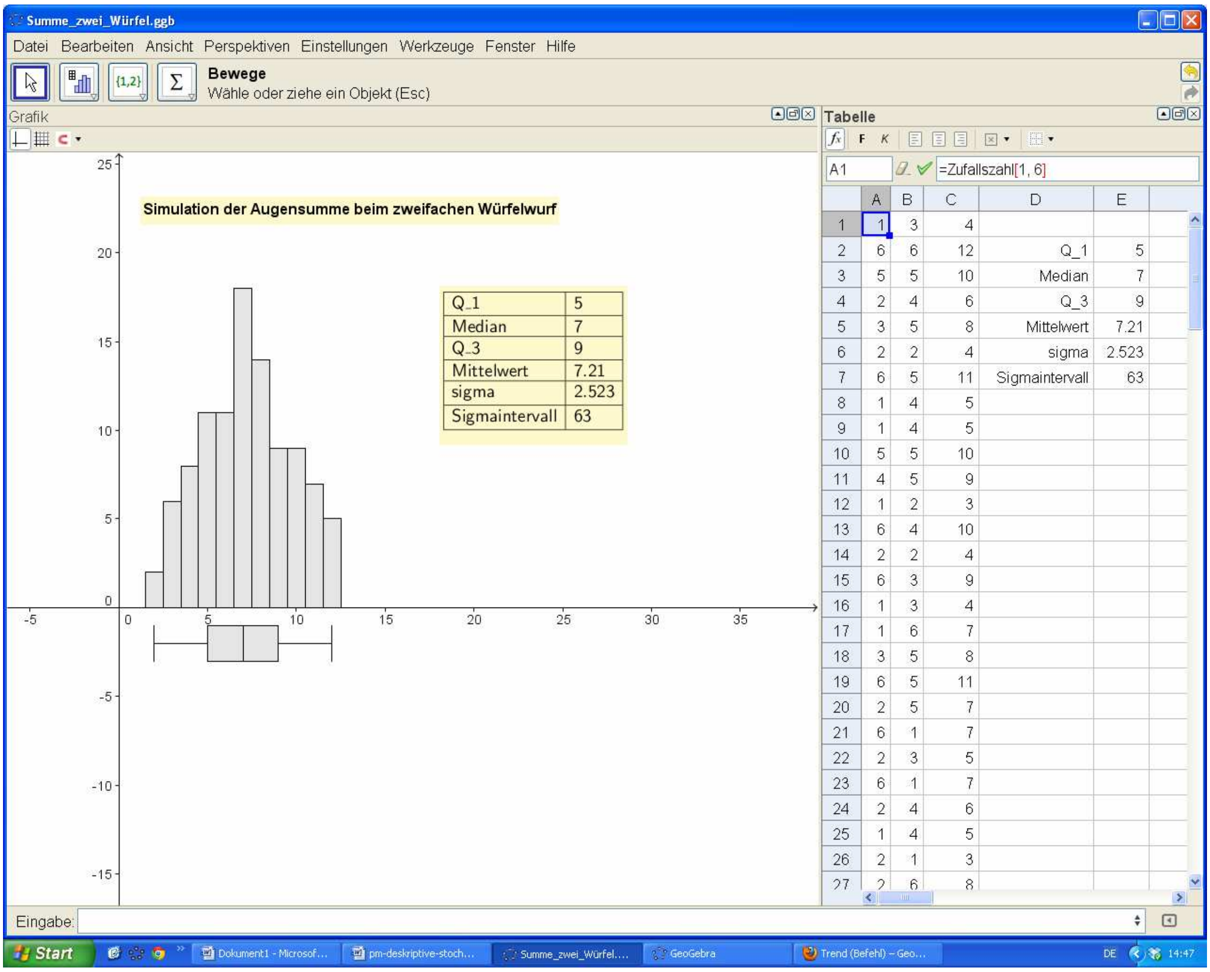


Abb.6: Verteilung der Augensumme bei 100- facher Wiederholung eines zweifachen Würfelwurfs

Wie man auch der Abb. 6 entnimmt, wurden in den Spalten A und B mit dem Befehl „Zufallszahl[1,6]“ jeweils 100 Zufallszahlen erzeugt und in der Spalte C wurde die jeweilige Summe berechnet. Die (mit ihrer Dreiecksform) schon „fast“ glockenförmige Verteilung wundert nicht, wenn man daran denkt, dass beispielsweise die Augensumme 7 auf 6 verschiedene Arten erzeugt werden kann, während die Augensumme 2 beispielsweise nur bei (1,1) zustande kommt. Schwieriger- aber auch interessanter- wird die Untersuchung der Augensummenhäufigkeit, wenn man Augensummen von größeren Wurffolgen betrachten möchte, vgl. Strick (1998). Die Simulation hingegen ist wieder leicht zu erzeugen, und erlaubt eine Bestätigung der Additivität der Standardabweichung: Man löscht die Inhalte der Spalten A und B und trägt stattdessen in Spalte C gleich die Augensumme einer gewünschten Zahl von Würfen ein, d.h. man schreibt in die Eingabezeile: C1=Summe[Folge[Zufallszahl[1, 6], k, 1, Wurfzahl]], wobei durch einen Schieberegler für natürliche Zahlen die Wurfzahl eingestellt wird. Eine angeleitete Erkundung zur Additivität der Standardabweichung findet sich in der Kopiervorlage am Ende es Artikels.

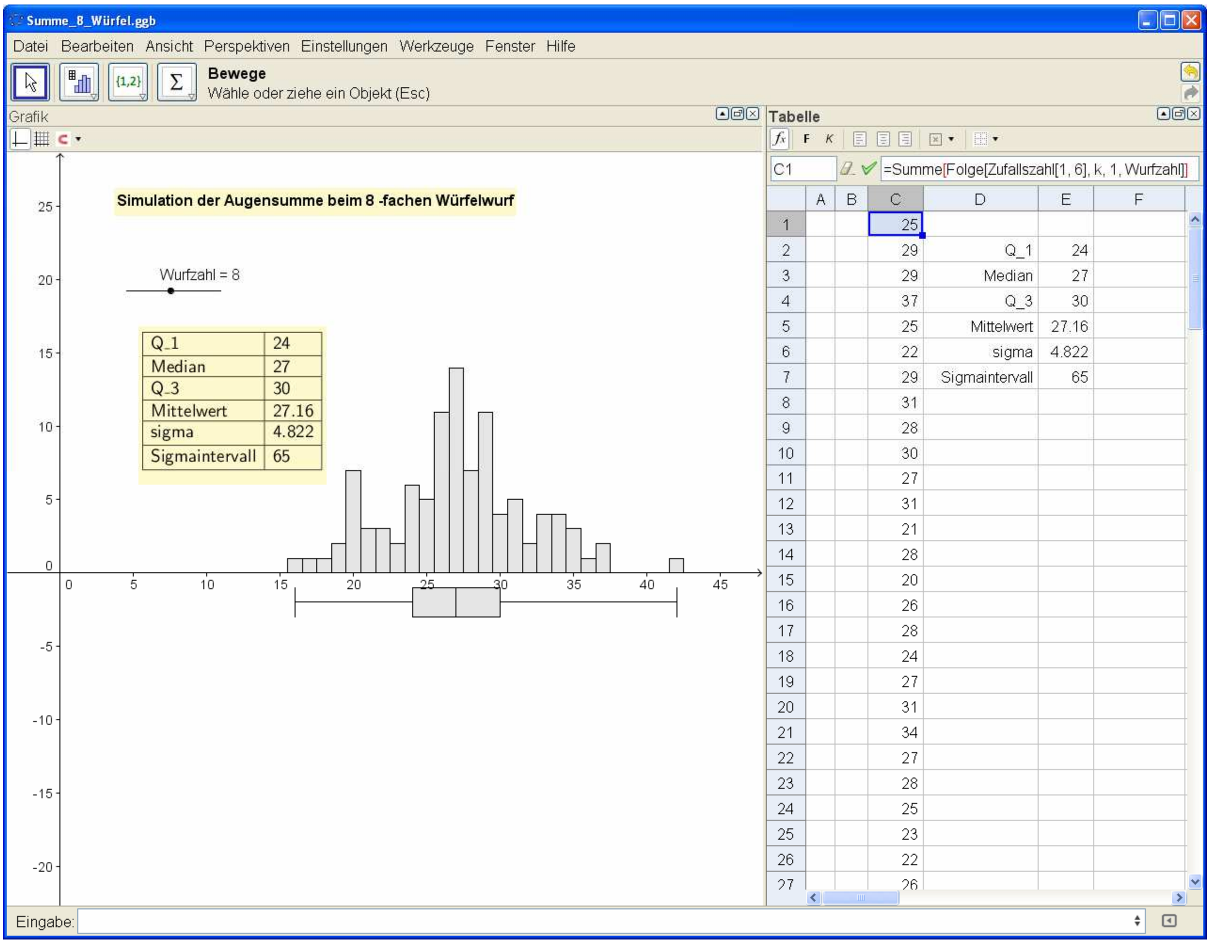


Abb.7: Verteilung der Augensumme bei 100- facher Wiederholung eines 8-fachen Würfelwurfs

Resümee

- Geogebra besitzt mit seiner Tabellenkalkulation und intuitiven Befehlen zur Visualisierung großer Datensätze Vorteile gegenüber klassischen Tabellenkalkulationen.
- Simulationen können tiefer liegende theoretische Zusammenhänge wie die Additivität der Varianz und den zentralen Grenzwertsatz im Rahmen beschreibender Statistik einsichtig machen.
- Wer darüber hinaus schon in der Sekundarstufe I beschreibende mit beurteilender Statistik verbinden möchte, wer also „knackige Probleme oder spannende Hypothesen sucht, die sich mit beschreibender Statistik und Simulationen beantworten lassen, der findet in Riemer/Seebach (2011) und Riemer (2012) weiter gehende Anregungen, die mit den hier vorgestellten Werkzeugen GeoGebras hervorragend umgesetzt werden können.

Literatur

Riemer, Wolfgang /Seebach, Günter (2011): Bleistiftrollen – Beurteilende Statistik im Federmäppchen. Wie zwischen Experimentieren und Simulieren Grundgedanken beurteilender Statistik reifen. In: Kaenders, Rainer /Schmidt, Rainer: Mit GeoGebra mehr Mathematik verstehen. Vieweg, Wiesbaden. 2011.

Riemer, Wolfgang (2012): Statistik mit Red Bull - Pharmaforschung im Klassenraum. mathematiklehren 6/2012.

Strick, Hein-Klaus (1998): Augensummen. Praxis der Mathematik, Heft 2 / Jg. 40

Dr. Wolfgang Riemer
Zentrum für Schulpraktische Lehrerausbildung (Studienseminar) Köln
August-Bebel-Str. 80, 50259 Pulheim
w.riemer@arcor.de

Günter Seebach
Im alten Garten 13
53773 Hennef
guenter.seebach@t-online.de

Standardabweichung und Gauss'sche Glocke **Kopiervorlage**

In dieser Erkundung soll die folgende Aussage experimentell untermauert werden: Wenn man viele unabhängige Daten X zu einer Summe S addiert, wird deren Verteilung glockenförmig und im Standardabweichungsintervall um den Mittelwert liegen ca. 68% aller Summenwerte.

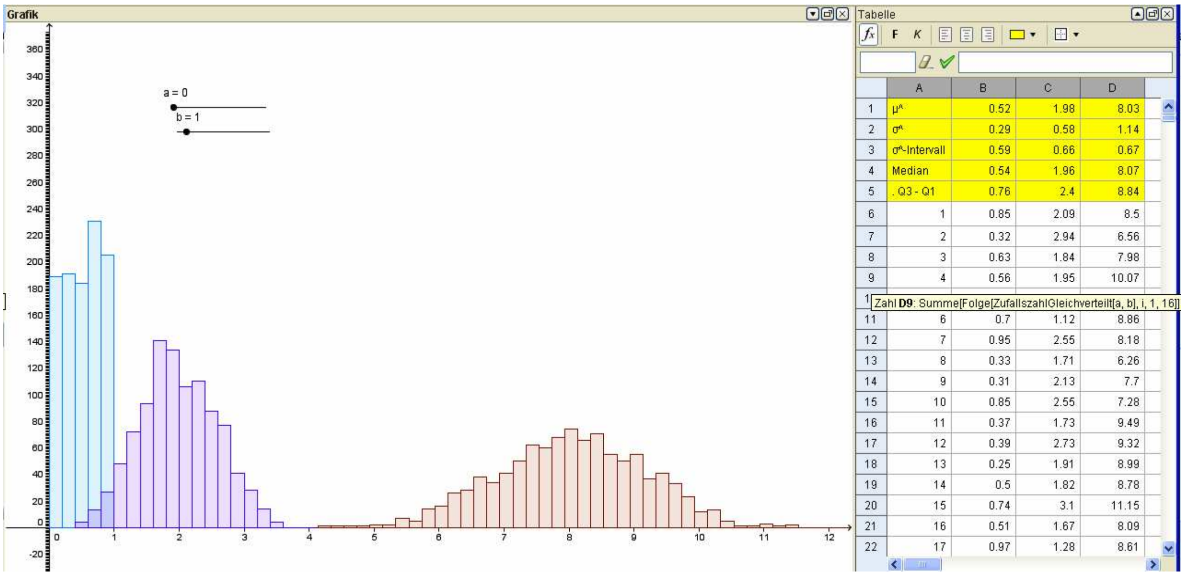


Abb. 8 Summen gleich verteilter Zufallszahlen

- Erzeuge mit dem Befehl ZufallszahlGleichverteilt[a,b] in der Spalte b von GeoGebras Tabellenkalkulation 1000 Zufallszahlen X. Wähle zunächst a=0 und b=1 (vgl. Abb. 7).
- Bestimme den Mittelwert μ^, die Standardabweichung σ^, den Median und den Quartilabstand Q3-Q1.

  Bestätige: μ^ schwankt um μ=0,5, σ^ schwankt zufallsabhängig um $\sigma = \sqrt{\frac{1}{12}} \approx 0{,}29$

  der Median schwankt um 0,5, der Quartilabstand schwankt zufallsabhängig um 0,5.
  Versuche, einige dieser Beobachtungen zu begründen.
- Schätze welcher Anteil der 1000 Zufallszahlen im Intervall $[\mu\hat{}-\sigma\hat{}\,;\mu\hat{}+\sigma\hat{}\,]$ liegt. Notiere Zahlenbeispiele.
- Ergänze die Tabelle durch zwei Spalten, in die du statt einzelner Zufallszahlen X Summen aus 4 bzw. 16 Zufallszahlen einträgst. Das kann durch Addieren: ZufallszahlGleichverteilt[a,b] + ... + ZufallszahlGleichverteilt[a,b] oder mit dem Befehl Summe[Folge[ZufallszahlGleichverteilt[a,b],,i,1,16]] geschehen (vgl. Abb. 8).
- Belege, dass die Verteilungen mit wachsender Summandenzahl immer glockenförmiger werden.
- Belege durch Notieren mehrerer Zahlenbeispiele, dass sich die Standardabweichungen, nicht aber die Quartilabstände bei Vervierfachung der Summandenzahl verdoppeln und dass die 68%-Sigmaregel für glockenförmige Summen-Verteilungen brauchbare Ergebnisse liefert.
- Mit ein wenig Neugier kannst du herausfinden, welche Zahlen du in GeoGebras Befehl ZufallszahlNormal[a,b] für a und b einsetzen musst, damit glockenförmige Verteilungen entstehen wie oben bei den Summen aus 4 bzw. 16 Summanden.