

Ausführliche Grundlagen zum Versuch 303

Standardabweichung

Wie entsteht die Standardabweichung? Bei *wissenschaftlichen Messreihen* resultiert sie aus dem Zusammenwirken aller Messfehler, d.h. die Standardabweichung s ist ein Maß für die Qualität der Messung. Sie sollte aber nicht als Fehler des Mittelwertes betrachtet werden, denn dieser ist in der Regel deutlich kleiner.

Bei von Natur aus *statistisch verteilten Messgrößen* ist eine gewisse Streuung von vorneherein gegeben. Sie hat nichts mit „Qualität der Messung“ zu tun, allenfalls mit der Messzeit (siehe unten). Im vorliegenden Fall einer *Poisson-Verteilung* steht die Standardabweichung in direktem Verhältnis zum Mittelwert. Sie ist für eine hinreichend große Zahl n von Messwerten gleich der Wurzel des Mittelwertes ($s = \sqrt{\bar{x}}$). Entsprechend gilt für die Varianz $s^2 = \bar{x}$.

Beim *Würfeln*, wo im Falle eines ideal gebauten Würfels und einer großen Zahl von Würfeln alle Zahlen von 1 bis 6 gleich oft vorkommen sollten, ist natürlich der Mittelwert (ebenso wie die immer gleiche Streuung) bereits vorher bekannt und ausrechenbar.

Bemerkung zu Gl.2 in Anleitung 303: Im Nenner der Gleichung ist durch den Term $(n-1)$ berücksichtigt, dass zur Schätzung der Standardabweichung schon ein anderer Schätzwert benutzt wird, nämlich der Mittelwert. Kennt man den wahren Wert, so wird durch n geteilt. In Statistiklehrbüchern findet man diesen Unterschied unter den Begriffen „Standardabweichung der Stichprobe“ ($n-1$) bzw. „Standardabweichung der Grundgesamtheit“ (n). Bei gängigen Softwareprodukten und modernen Taschenrechnern können meist beide Werte berechnet werden.

Genauigkeit des Mittelwertes

Bei physikalischen Messreihen (also auch im Praktikum) interessiert neben der Größe des Mittelwertes \bar{x} vor allem dessen Genauigkeit $\Delta\bar{x}$ (Stichwort: Fehlerrechnung). Oft wird für die Genauigkeit des Mittelwertes die Standardabweichung verwendet. Diese auf den ersten Blick naheliegende Vermutung ($\Delta\bar{x} = \pm s$) erweist sich jedoch als falsch, denn wenn man die Messreihe mehrmals hintereinander wiederholt und jeweils den Mittelwert berechnet, dann sieht man, dass die Streuung $s_{\bar{x}}$ der Mittelwerte viel kleiner ist als die Streuung s der einzelnen Messwerte. Das liegt daran, dass sich durch die Mittelwertbildung die großen „Ausreißer“ (nach oben und unten) gegenseitig aufheben. Die Genauigkeit $\Delta\bar{x}$ (Streubreite der Mittelwerte $s_{\bar{x}}$) ist $\Delta\bar{x} = \pm s/\sqrt{n}$. Im Versuch wird dieser Zusammenhang mit Hilfe des Würfel-Simulationsprogramms nachgewiesen (er lässt sich auch auf Basis statistischer Axiome herleiten).

Durch die Aufnahme vieler Messwerte verbessert man also nicht die Güte der Messung an sich (s ist unabhängig von n), dafür aber die Zuverlässigkeit der Schätzung von \bar{x} ($\Delta\bar{x}$ wird kleiner). Man könnte allerdings auch s beeinflussen (durch genaueres Messen).

Bei rein statistischen Messungen ist es anders. Dort wird die Genauigkeit des Mittelwertes allein durch die Anzahl der Messwerte (meistens limitiert durch die Messzeit) bestimmt. Der Abschnitt „Wann wird aus der Poisson- eine Normalverteilung?“ zeigt anhand eines Beispiels, wie sich durch Vergrößerung der Messzeit der relative Fehler verringert.

Normalverteilung

Die Normalverteilung (Gaußsche Glockenkurve) ist die bekannteste und wichtigste statistische Verteilung. Das liegt zum einen daran, dass sie als mathematische Funktion analytisch einfacher zu handhaben ist als diskrete Zusammenhänge wie Zählraten und Ziffern und es daher schon

aus praktischen Gründen erstrebenswert ist, eine gegebene Verteilung durch eine Normalverteilung anzunähern. Ihre herausragende Bedeutung erhält sie aber durch die Gültigkeit des *Zentralen Grenzwertsatzes*, womit sie bei vielen Messungen in Naturwissenschaft und Technik als die Fehlerverteilung schlechthin betrachtet werden kann. Eine Normalverteilung entsteht immer dann, wenn sich viele unabhängige Zufallsgrößen aufsummieren, was in der Praxis häufig vorkommt, aber (Vorsicht!) im konkreten Fall ggf. auch erstmal nachgewiesen werden muss (nicht alles ist normalverteilt).

Bei der Gaußkurve liegen ca. 68% aller Werte im Intervall $\mu \pm \sigma$ (1σ -Bereich) um den Mittelwert μ , ca. 95% im Intervall $\mu \pm 2\sigma$ (2σ -Bereich), ca. 99.7% im Intervall $\mu \pm 3\sigma$ (3σ -Bereich) usw. Die Gleichung der Wahrscheinlichkeitsdichteverteilung $w(x)$ lautet:

$$w(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Bemerkung zur Verwendung der Variablen \bar{x} , s , μ und σ : Bei der Normalverteilung wird der Zentralwert mit μ und die Standardabweichung mit σ bezeichnet. Diese Bezeichnungsweise bietet sich allgemein für theoretische Verteilungen bzw. (unendlich große) Grundgesamtheiten an. Bei endlichen Stichproben ist es dagegen besser, \bar{x} und s zu verwenden.

Experiment (Stichprobe)	Theorie (Grundgesamtheit)
\bar{x}	μ
s	σ

Zentraler Grenzwertsatz der Statistik

Der Zentrale Grenzwertsatz besagt, dass Summen von jeweils n Zufallszahlen, die derselben Grundgesamtheit entnommen werden, für große n eine Normalverteilung bilden, unabhängig davon, wie die einzelnen Zufallszahlen selber verteilt sind. Ursache ist das schon mehrfach erwähnte gegenseitige Kompensieren der Ausreißer, und es gilt wieder: je größer die Anzahl der Summanden, desto deutlicher ist der Effekt.

Teilt man eine Summe durch die Anzahl ihrer Summanden n , so erhält man das arithmetische Mittel. Damit kann der Satz in gleicher Weise auch für die Mittelwerte formuliert werden: Die *Mittelwerte* \bar{x} von Stichproben, bestehend aus jeweils n Zufallszahlen x_k einer beliebiger Ausgangsverteilung, sind immer normalverteilt. Dabei ist der Mittelwert dieser Verteilung derselbe wie der von x_k , nämlich \bar{x} , während die Varianz $s_{\bar{x}}^2$ um den Faktor n kleiner ist als die Varianz s^2 der Ausgangsverteilung. Für die Standardabweichung $s_{\bar{x}}$ der Mittelwert-Verteilung gilt daher: $s_{\bar{x}} = s/\sqrt{n}$.

Poisson-Verteilung

Die Zählraten der natürlichen Radioaktivität bilden eine Poisson-Verteilung. Die Poisson-Verteilung ist in der Natur sehr häufig anzutreffen. Sie entsteht immer dann, wenn eine sehr geringe Ereigniswahrscheinlichkeit (radioaktives Isotop mit einer Halbwertszeit von Millionen Jahren zerfällt ausgerechnet jetzt, während wir es beobachten) einer sehr großen Zahl von Ereignismöglichkeiten (es gibt extrem viele dieser Isotope) gegenübersteht.

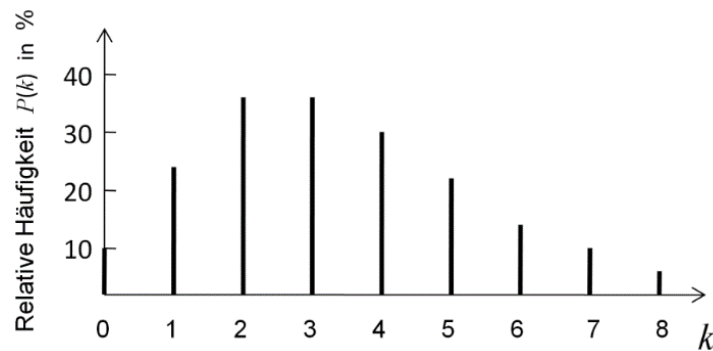


Bild 1: Poisson-Verteilung mit $\bar{x} = 3$.

Die zugehörige Gleichung lautet:

$$P(k) = \lim_{\substack{p \rightarrow 0 \\ n \rightarrow \infty}} B(k, n, p) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \quad \text{mit} \quad \bar{x} = s^2 = \lambda$$

Die Herleitung der Formel (Grenzübergang) kann Statistik-Lehrbüchern entnommen werden. Die Poisson-Verteilung hat eine gewisse Ähnlichkeit mit der Gaußschen Glockenkurve und kann für große Mittelwerte \bar{x} auch durch diese angenähert werden (Erklärung siehe unten). Für kleine Mittelwerte ist sie aber deutlich unsymmetrisch, sie besitzt eine sog. „*Schiefe*“. Dass eine Normalverteilung ursprünglich aus einer Poisson-Verteilung hervorgegangen ist, sieht man am Zusammenhang $s^2 = \bar{x}$ bzw. $s = \sqrt{\bar{x}}$, d.h. wenn man eine Normalverteilung findet, deren Breite s etwa gleich der Wurzel ihres Mittelwertes ist, so kann man vermuten, dass ihr eigentlich ein poissonverteilter Zusammenhang zugrunde liegt.

Wann wird aus der Poisson- eine Normalverteilung?

Im Versuch wird mit einer relativ kleinen Zählzeit von ca. 2s gearbeitet. Damit erhalten wir Werte im Bereich 0 bis 10 unsymmetrisch verteilt mit einer Häufung bei 3. Würden wir die Zählzeit auf 200s stellen, kämen Werte im Bereich 250 bis 350 heraus, jetzt aber nahezu symmetrisch um den Mittelwert 300. Warum? Durch die Erhöhung auf $100 \cdot 2s = 200s$ werden praktisch 100 Zufallsgrößen im Bereich 0 ... 10 intern aufaddiert und erst am Ende die Summe abgefragt. Statistisch betrachtet sind unsere Messwerte in diesem Fall „Summen von unabhängigen Zufallsgrößen derselben Grundgesamtheit“ (genau wie vom Zentralen Grenzwertsatz verlangt), d.h. jetzt sollten die gemessenen Zählraten näherungsweise normalverteilt sein. Die ursprüngliche Poissonverteilung ist aber noch daran zu erkennen, dass die Standardabweichung etwa der Wurzel des Mittelwertes entspricht. Ganz nebenbei (aber für die Praxis relevant) erhöht sich mit der Messzeit die (relative) Genauigkeit der Messung, die Verteilung wird bezogen auf den Mittelwert schmaler. Bei 2s ist $\bar{x} = 3$ und damit $s \approx 1.73$, bei 200s wäre $\bar{x} = 300$ und $s \approx 17.3$; der relative „Fehler“ (s/\bar{x}) ist im ersten Fall 58%, im zweiten Fall nur noch 5.8%. Die Erhöhung der Messzeit um den Faktor 100 verbessert also die Messgenauigkeit um den Faktor $\sqrt{100} = 10$. Dasselbe wäre passiert, hätte man die Messung mit 2s Zählzeit 100 mal wiederholt und dann statistisch ausgewertet.

Statistik und Fehlerrechnung

Bei der Messung physikalischer Größen existiert im Gegensatz zu den im vorliegenden Experiment untersuchten Beispielen (Würfeln, radioaktiver Zerfall) nicht ein „wahrscheinlichster“ sondern tatsächlich ein „wahrer“ Wert. Dass dieser mit einer einmaligen Messung nicht genau getroffen wird, liegt an der begrenzten Messgenauigkeit, verursacht durch die zufälligen Fehler. Es ist sinnvoll, immer dort wo es möglich ist, eine Messung mehrmals zu wiederholen. Das hat mehrere Vorteile:

- 1) Es lassen sich grobe Fehler erkennen.
- 2) Man erhält ein Gefühl für die erreichbare Genauigkeit.
- 3) Die Genauigkeit des Ergebnisses wird durch Mittelwertbildung besser.

Durch das Wiederholen der Messung erhalten wir eine stichprobenartige Folge (unabhängiger) Messwerte, d.h. eine Verteilung von n Einzelwerten x_k , von der wir das arithmetische Mittel \bar{x} und die Standardabweichung s berechnen können. Die Standardabweichung ist, wie bereits oben geschrieben, zwar ein gutes Maß für unsere Genauigkeit, aber sie ist noch nicht der Fehler des Mittelwertes, was man dann sieht, wenn dieselbe Messreihe mehrfach wiederholt wird. Dann ist die Streuung der Mittelwerte nämlich deutlich kleiner (um \sqrt{n} kleiner) als die Streuung s der einzelnen Messwerte.

Welchen Fehler soll man nun für den Mittelwert angeben und wie sicher ist dieser Wert? Das folgende Gedankenexperiment liefert die Antwort auf diese Frage: Ausgehend vom Zentralen Grenzwertsatz kann man den Mittelwert unserer Messreihe als Teil einer Normalverteilung betrachten, die entstanden wäre, hätten wir die Messreihe mehrmals (z.B. tausend Mal, so wie bei der Würfelsimulation) wiederholt. Deren Maximum entspräche dann ziemlich gut dem wahren Wert. Unser Mittelwert könnte zufällig in der Nähe des Maximums liegen, aber auch zufällig ganz weit außen, wir wissen es nicht (da in der Praxis natürlich keine 1000 Wiederholungen gemacht werden). Was wir aber wissen (da wir die Eigenschaften der Gaußschen Glockenkurve kennen) ist, dass er mit 68% Wahrscheinlichkeit im Bereich $\pm s_{\bar{x}}$ um den wahren Wert (entsprechend $\mu \pm \sigma$) liegt, da ein Intervall der Breite $\pm s_{\bar{x}}$ um den wahren Wert 68% aller möglichen Mittelwerte enthalten würde, die beim mehrmaligen Wiederholen der Messreihe aufgetreten wären. D.h. wir können für unser Ergebnis \bar{x} als Fehler ruhig den Wert $\Delta\bar{x} = s_{\bar{x}} = \pm s/\sqrt{n}$ angeben, müssen aber hinzufügen, dass diese Genauigkeit nur in ca. 68% aller Fälle garantiert werden kann.

Die Genauigkeit (*Vertrauensbereich, Konfidenzintervall*) eines statistischen Resultats kann also immer nur mit einer begrenzten *statistischen Sicherheit angegeben* werden. Das muss man akzeptieren.

Die Fehlerangabe $\bar{x} \pm s/\sqrt{n}$ bietet eine 68%-ige Wahrscheinlichkeit dafür, dass der wahre Wert im Intervall $(\bar{x} - s/\sqrt{n})$ bis $(\bar{x} + s/\sqrt{n})$ liegt (*Vertrauensbereich für 68% statistische Sicherheit*). Das ist nicht schlecht, bedeutet aber auch, dass der von uns angegebene Fehler in einem Drittel der Fälle zu klein wäre. Würde man das Fehlerintervall verdoppeln: $\Delta x = 2 \cdot s/\sqrt{n}$ (entsprechend $\mu \pm 2\sigma$), dann würde die Sicherheit der Genauigkeitsangabe schon 95% betragen. Da in der Statistik keine 100%-igen Aussagen möglich sind, geht es immer darum, unter Berücksichtigung der gewünschten bzw. geforderten statistischen Sicherheit den zugehörigen „Fehler“ bzw. besser „Vertrauensbereich“ $(\bar{x} \pm t \cdot s/\sqrt{n})$ des Ergebnisses \bar{x} anzugeben.

***t*-Verteilung**

Die Größe t , die im letzten Satz des vorhergehenden Abschnitts aufgetaucht ist (der sogenannte „ t -Parameter“), kommt von der „Student- oder t -Verteilung“ (vgl. Literatur) und kann ganzzahlig (1, 2, 3, ...) aber auch beliebig „krumm“ sein. Sie hängt von der Statistischen Sicherheit p bzw. der Irrtumswahrscheinlichkeit $\alpha = (1 - p)$ und vom Stichprobenumfang n ab und kann Tabellen entnommen werden.

Tab. 1: Parameter t in Abhängigkeit von der Wahrscheinlichkeit p und der Stichprobengröße n .

p	$n = 5$	$n = 10$	$n = 20$	$n = 50$	$n = 100$
68%	1.08	1.05	1.02	1.01	1.00
95%	2.78	2.26	2.09	2.00	1.98
99%	4.60	3.25	2.85	2.68	2.64

Auch bei der **Linearen Regression** muss die statistische Sicherheit beachtet werden. Die üblichen Computerprogramme geben als Fehler für Anstieg und Achsenschnittpunkt den Standardfehler aus. Dieser gilt nur für 68% aller Fälle (und auch nur für große n). Bei einer gewünschten statistischen Sicherheit von 95% muss noch mit dem zugehörigen t - Parameter (vgl. Tabelle) multipliziert werden.

Tab.2: Parameter t bei linearer Regression für $p = 95\%$ in Abhängigkeit von der Zahl der Messpunkte.

Anzahl der Messpunkte	5	10	20	100
t - Parameter	3.2	2.3	2.1	2.0