

Normalverteilung | MatheGuru

Eine der wichtigsten stetigen Wahrscheinlichkeitsverteilungen, ist die Normalverteilung. Sie wurde von Abraham de Moivre und später Carl Friedrich Gauß analysiert. Der Beitrag von Gauß war so fundamental, dass die Normalverteilung auch oft **Gauß-Verteilung** genannt wird. Wegen ihrer charakteristischen Form wird sie auch manchmal einfach nur **Glockenkurve** genannt, auch wenn es viele Verteilungsfunktionen gibt, die einen glockenförmigen Graphen besitzen.

Die Einsatzmöglichkeiten der Normalverteilung sind so zahlreich, dass sie als das "Schweizer Taschenmesser" der Statistik bezeichnet werden kann.

Definition

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

Das Aussehen und die Eigenschaften der Normalverteilung werden durch zwei Parameter bestimmt:

- Der **Erwartungswert** μ . Er legt fest, an welcher Stelle die Normalverteilung ihr Maximum haben wird.
- Die **Varianz** σ^2 . Die Wurzel der Varianz σ ist die [Standardabweichung](#).

Die gesamte Fläche, die von der Kurve der Normalverteilung eingeschlossen wird (daher das Integral von $-\infty$ bis ∞), ist stets 1.

Anwendung

Intelligenz, Körpergröße (eines einzigen Geschlechts), sogar Sozialkompetenz: all diese Werte sind normalverteilt. Dies bedeutet beispielsweise, dass die meisten Menschen durchschnittlich groß sind und nur sehr wenige sehr groß oder sehr klein sind. Sogar Einkommen wird normalverteilt, wenn man die Daten vorher logarithmiert.

Die Normalverteilung ist die wichtigste Verteilung der Statistik, und wird sowohl in Naturwissenschaften als auch Geistes- und Wirtschaftswissenschaften verwendet, deren tatsächliche Verteilungsfunktion unbekannt ist. Sie wird meist verwendet, wenn die eigentliche, den Daten zugrunde liegende Verteilungsfunktion unbekannt ist. Ein Grund für den hohen Stellenwert der Normalverteilung ist der [zentrale Grenzwertsatz](#).

Der Zentrale Grenzwertsatz

Zum Hauptartikel: [Zentraler Grenzwertsatz](#).

Er besagt, dass der Durchschnitt einer großen Anzahl an beobachteten Zufallsvariablen, die aus derselben Verteilung gezogen wurden, annähernd normalverteilt sein werden, unabhängig von der Verteilungsfunktion aus der sie herausgenommen wurden. Es ist daher so, dass physische Quantitäten, welche die Summe aus vielen verschiedenen Unterprozessen sind (wie beispielsweise Messfehler) oft eine Verteilungsfunktion haben, die annähernd der Normalverteilung entspricht.

Nimmt man also eine ausreichend große Stichprobe aus einer Grundgesamtheit, so wird der Mittelwert der

Stichprobe näherungsweise dem Mittelwert der Grundgesamtheit entsprechen. Zusätzlich werden alle Stichproben annähernd normalverteilt sein, mit einer Varianz, die der der Grundgesamtheit geteilt durch die Größe der Stichprobe entspricht.

Ein weiterer Grund für die Beliebtheit der Normalverteilung ist, dass andere Größen analytisch hergeleitet werden können, wenn man sie als Verteilungsfunktion annimmt. Eine dieser Größen ist beispielsweise die **Fehlerfortpflanzung**.

Standardnormalverteilung

Der einfachste Fall tritt ein, wenn $\mu = 0$ und $\sigma^2 = 1$ ist. Für diese Werte wird die Normalverteilung auch **Standardnormalverteilung** genannt.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}$$

Der Vorfaktor $\frac{1}{\sqrt{2\pi}}$ stellt sicher, dass die gesamte Fläche unter der Kurve (und damit auch das Integral von $-\infty$ bis ∞) eine Fläche von genau 1 hat. Die $\frac{1}{2}$ im Exponenten der e-Funktion gibt der Normalverteilung eine Einheitsvarianz (und damit auch eine Einheitsstandardabweichung). Die Symmetrieachse der Funktion liegt

bei $x=0$, wo sie auch ihren Höchstwert $\frac{1}{\sqrt{2\pi}}$ erreicht. Die beiden Wendestellen liegen jeweils bei $x = 1$ und $x = -1$.

Allgemeine Normalverteilung

Jede Normalverteilung ist eine Variante der Standardnormalverteilung. Um die Funktionswerte anzugleichen, muss die Standardnormalverteilung auf zwei verschiedene Weisen verändert werden:

- Die Standardnormalverteilung muss um den Faktor $\frac{1}{\sigma}$ gestreckt werden. Dies stellt auch sicher, dass die vom Graphen eingeschlossene Fläche weiterhin 1 bleibt.
- Der Funktionsparameter der Standardnormalverteilung wird dabei **z-transformiert** (auch **z-standardisiert**). Diese Art der Transformation berechnet einen neuen Wert zu einer Normalverteilungsfunktion mit dem Erwartungswert Null und der Varianz Eins.

$$\underbrace{\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}_{\text{Allgemeine Normalverteilung}} = \frac{1}{\sigma} \cdot \underbrace{\phi\left(\frac{x-\mu}{\sigma}\right)}_{\text{Standardnormalverteilung}}$$

Schreibweise

Die Standardnormalverteilung wird in der Literatur oft mit dem kleinen griechischen Buchstaben ϕ (phi) angegeben. Die andere Schreibweise dieses Buchstaben φ wird auch verwendet.

Will man die Normalverteilung allerdings mit den Parametern für den Erwartungswert und der Varianz

angeben, schreibt man $\mathcal{N}(\mu, \sigma^2)$. Daher, wenn eine Zufallsvariable X mit einem Erwartungswert von μ und einer Varianz von σ^2 normalverteilt ist, schreibt man:

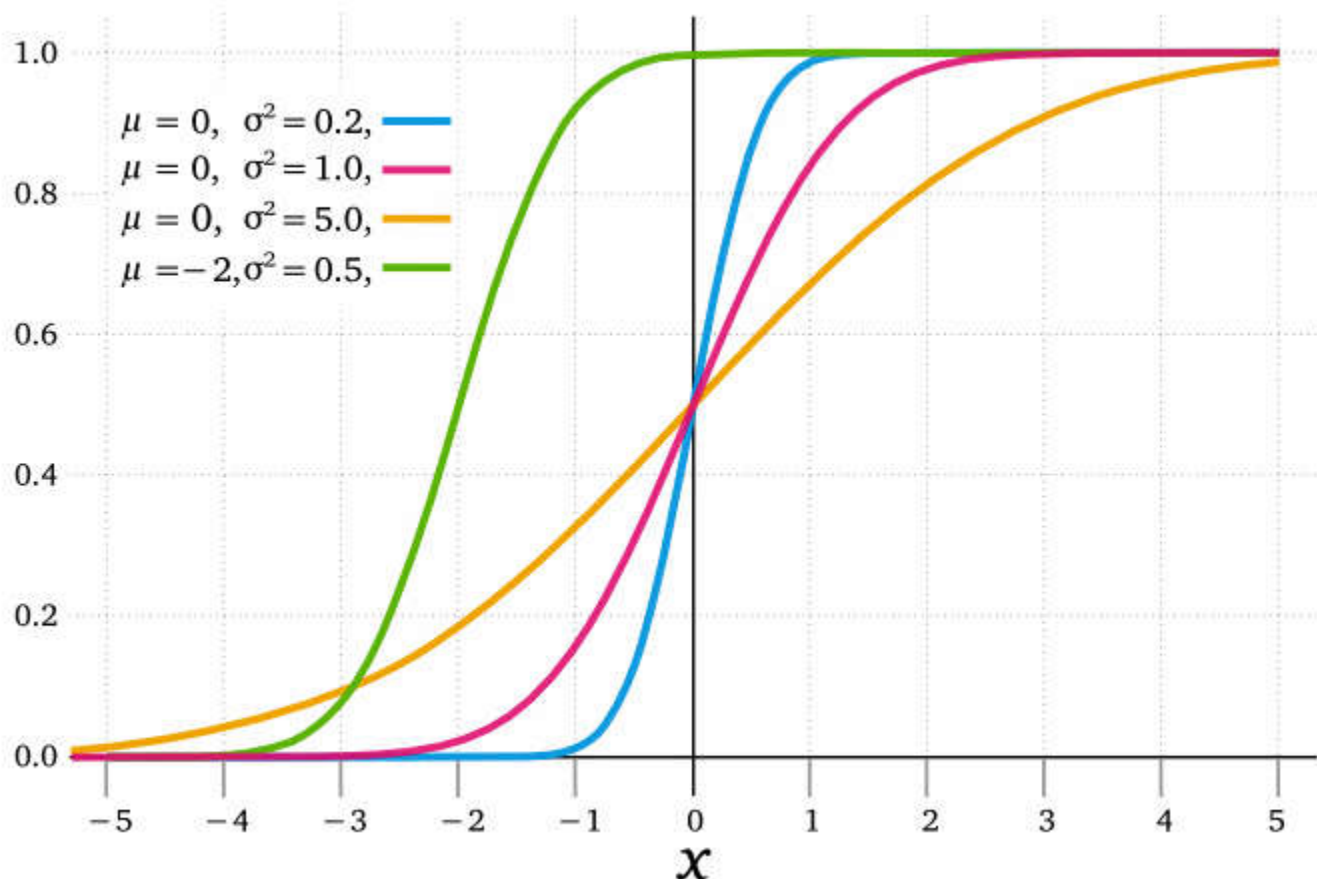
$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Verteilungsfunktion der Normalverteilung

Die Verteilungsfunktion der Normalfunktion ist die eingeschlossene Fläche unter der Normalfunktion (daher das Integral) von $-\infty$ bis zum Wert x an. Sie hat einen schwanenhalsförmigen (Sigmoid) Graphen.

$$\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right) \right]$$

$\Phi(x)$ ist das Symbol für die Verteilungsfunktion der Standardnormalverteilung. Unterhalb sind die Graphen von vier Verteilungsfunktionen von vier Normalverteilungen für verschiedene Werte von μ und σ .



$\operatorname{erf}(x)$ ist die Gauß'sche Fehlerfunktion. Sie gehört zu den speziellen Funktionen und lässt sich nur als unendliche Reihe oder Kettenbruch (welcher auch unendlich ist) darstellen (siehe Definition unten). Normalerweise wird sie mit dem Computer oder Taschenrechner mit bereits vordefinierten Funktionen berechnet, daher ist vertiefendes Wissen bezüglich ihrer Berechnung in der Regel nicht notwendig.

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n \cdot x^{2n+1}}{(2n+1) \cdot n!} = \frac{2}{\sqrt{\pi}} \left(x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \frac{x^9}{216} - \dots \right) = 1 - -$$

1

Eigenschaften

Wert	Eigenschaft
Notation	$\mathcal{N}(\mu, \sigma^2)$
Parameter	$\mu \in \mathbb{R}$ (Erwartungswert) $\sigma > 0$ (Varianz)
Dichtefunktion	$\frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Verteilungsfunktion	$\frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x - \mu}{\sqrt{2\sigma^2}} \right) \right]$
Mittelwert	μ
Median	μ
Modus	μ
Varianz	σ^2
Standardabweichung	σ

Die Normalverteilung ist symmetrisch, wobei $x = \mu$ die Symmetrieachse bildet. Auch wenn sich die Werte der Normalverteilung [asymptotisch](#) dem Wert Null (nach beiden Seiten hin) nähern, so ist die Normalverteilung für keinen Wert von x jemals 0.

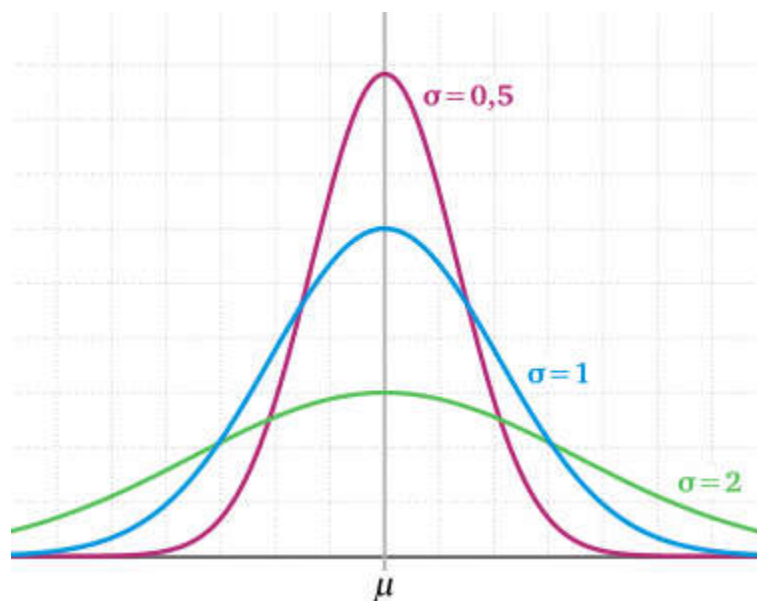
Die Normalverteilung erreicht auch Werte nahe Null, für Werte von x , die einige Standardabweichungen vom Erwartungswert entfernt liegen. Daher ist sie nicht unbedingt die Verteilungsfunktion der Wahl, wenn man eine größere Anzahl an Ausreißern (Werte, die einige Standardabweichungen vom Erwartungswert entfernt liegen) erwartet. Die Methode der kleinsten Quadrate und anderen Methoden der statistischen Inferenz, die sich optimal für normalverteilte Variablen anwenden lassen, geben in solchen Fällen nur sehr unzuverlässige Ergebnisse. Ist dies der Fall, sollten endlastige Verteilungen (*Heavy-tailed-Verteilung*) stattdessen verwendet werden.

Die Form der Dichtefunktion wird vollständig über die Standardabweichung σ bestimmt. Je kleiner σ ist, desto steiler ist der Gipfel der Funktion um den Erwartungswert herum; je größer σ , desto flacher ist der Graph.

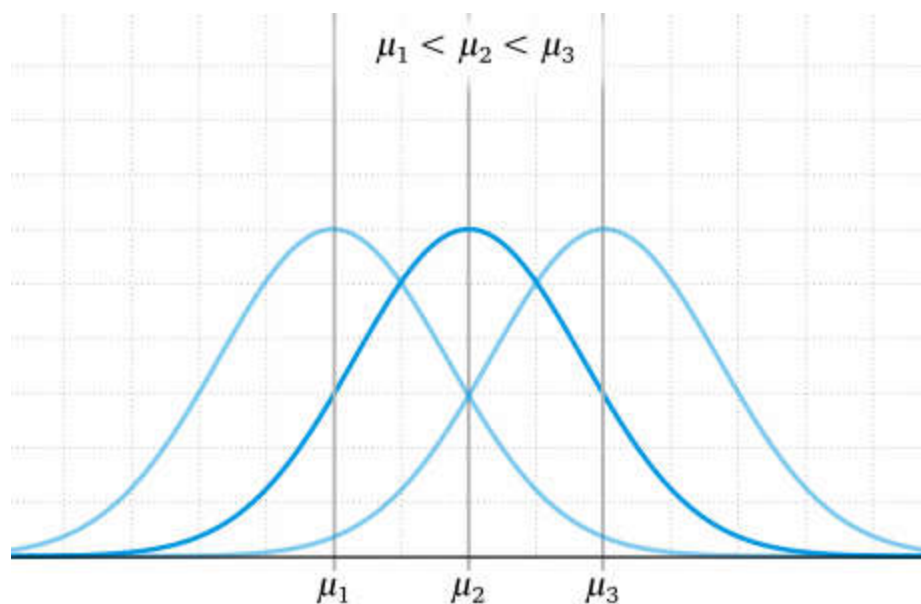
Der Parameter μ hingegen verschiebt die Normalverteilung entlang der x -Achse. Dies ist auch selbsterklärend, da die Normalverteilung immer bei μ ihr [Maximum](#) hat.

Die Veränderungen von der Standardabweichung σ und dem Erwartungswert μ und ihre Auswirkungen auf

den Graphen der Normalverteilung sind noch einmal in den Grafiken unterhalb zusammengefasst:

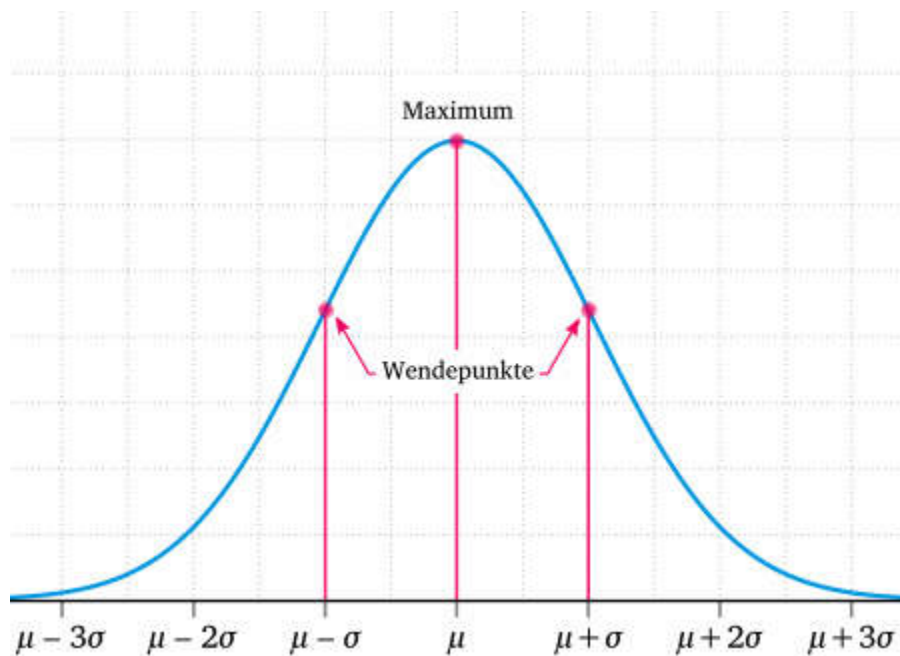


Kleinere Werte von σ lassen den Graphen der Normalverteilung um den Erwartungswert herum steiler verlaufen.



An der Stelle des Erwartungswerts μ ist das Maximum der Normalverteilungsfunktion. Ist der Wert kleiner, so bewegt sich der Graph weiter nach links; ist er größer, so verschiebt er sich nach rechts.

Weitere Eigenschaften



Eine Normalverteilung mit einem Erwartungswert und einer beliebigen Standardabweichung σ hat die folgenden Eigenschaften:

- Sie ist **symmetrisch**, wobei die vertikale Achse der Symmetrie bei $x = \mu$ liegt, welche auch der Modus, Median und Erwartungswert der Verteilung ist.
- Sie ist **unimodal** (sie hat nur einen Gipfel).
- Sie erreicht ihr **Maximum** an der Stelle $x = \mu$.
- Ihre **erste Ableitung** ist positiv für Werte von $x < \mu$ und negativ für Werte von $x > \mu$; am Punkt $x = \mu$ hat die erste Ableitung einen Wert von Null.
- Sie hat genau **zwei Wendestellen**: beide Wendestellen sind genau eine Standardabweichung vom Erwartungswert entfernt, nämlich bei $x_1 = \mu - \sigma$ und $x_2 = \mu + \sigma$.
- Sie ist an jeder Stelle von x **differenzierbar**.
- Sie ist **stetig**, daher von $-\infty$ bis ∞ definiert.

Spezielle Eigenschaften der Standardnormalverteilung

Die Standardnormalverteilung, als besondere Variante der Normalverteilung, hat zusätzlich noch folgende Eigenschaften:

- Ihre erste Ableitung $\phi'(x)$ ist gleich $-x \cdot \phi(x)$
- Die zweite Ableitung $\phi''(x)$ ist gleich $(x^2 - 1) \cdot \phi(x)$

Die Normalverteilung als Näherung zur Binomialverteilung

Wenn n ausreichend groß ist (etwa $n > 20$), ist die [Schiefe der Verteilung](#) klein genug, dass die **Normalverteilung zur Approximation der Binomialverteilung** $B(n, p)$ verwendet werden kann. In diesem Fall wird $\mathcal{N}(n, p, np(1-p))$ für die Parameter der Normalverteilung verwendet.

Im allgemeinen nähert die Normalverteilung die Binomialverteilung umso besser, umso größer n ist. Gleichzeitig sollte p nicht in der Nähe von 0 oder 1 sein – daher nahe 0,5. Es gibt eine Reihe von Faustregeln, die helfen eine Aussage darüber zu treffen, ob n und p adäquat gewählt wurden, um die Normalverteilung als Näherung zu verwenden:

- eine Faustregel besagt, dass $n \cdot p$ und $n(1-p)$ jeweils größer als 5 sein müssen. Jedoch gibt es auch Quellen, die 4 oder andere Zahlen als Mindestwert angeben. Generell kommt es auch darauf an, wie gut die Approximation sein soll. Daher gibt es auch Quellen, welche als Mindestwert 10 angeben – ein Wert welcher so hoch ist, dass gewisse Konvergierungseffekte eintreten. Dass heißt, dass für $n \rightarrow \infty$ die Werte der Binomialverteilung, den Werten die Normalverteilung entsprechen werden.
- Eine weitere Faustregel besagt, dass die Normalverteilung zur Näherung der Binomialverteilung verwendet werden darf, wenn $n > 5$ und

$$\left| \frac{1}{\sqrt{n}} \cdot \left(\sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}} \right) \right|$$

68–95–99,7-Regel

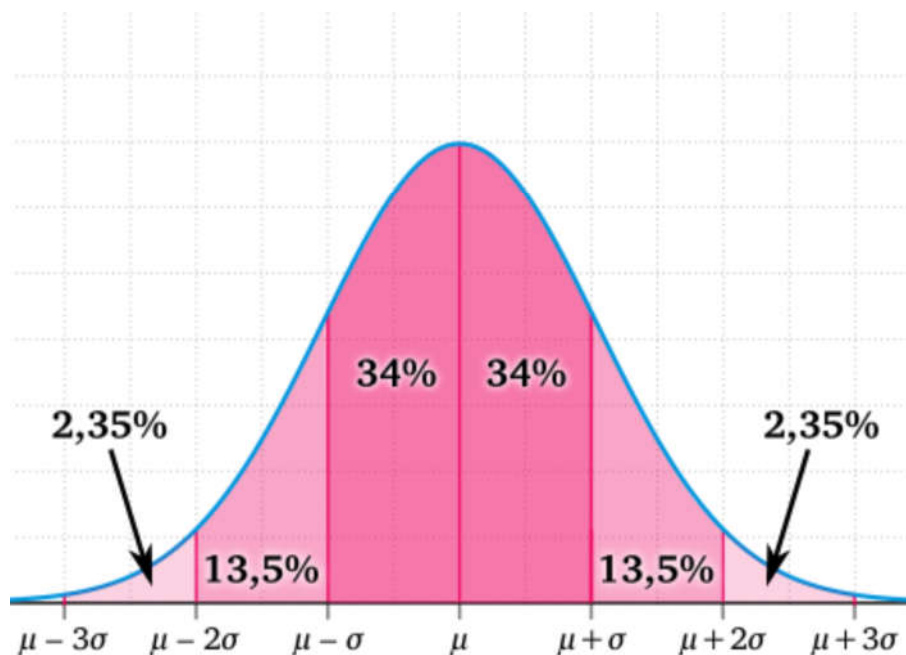
$$P(\mu - \sigma \leq x \leq \mu + \sigma) \approx 0,6827$$

$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) \approx 0,9545$$

$P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) \approx 0,9973$ Die 68-95-99,7-Regel gibt an, dass bei einer Normalverteilung **fast alle Werte innerhalb drei Standardabweichungen** vom Mittelwert aus fallen. Ungefähr 68,27% der Werte liegen innerhalb einer Standardabweichung vom Mittelwert. Ebenso liegen ungefähr 95,45% der Werte innerhalb von zwei Standardabweichungen vom Mittelwert. Und ca. 99,73% der Werte befinden sich innerhalb von drei Standardabweichungen vom Mittelwert.

Diese Regel gilt für alle Normalverteilungen – unabhängig vom Erwartungswert und der Standardabweichung.

Beispiel



Die Körpergröße des Menschen ist für ein Geschlecht betrachtet normalverteilt. Laut einer Statistik des sozio-oekonomischen Panels (SOEP) aus dem Jahr 2006 liegt der Erwartungswert der Durchschnittsgröße μ bei Frauen in Deutschland bei 165,4cm, und die Standardabweichung σ bei 4,5cm.

Aus der 68-95-99,7-Regel folgt, dass

- **68%** aller deutschen Frauen eine Körpergröße zwischen 160,9cm ($\mu - \sigma$) und 169,9cm ($\mu + \sigma$) haben
- **95%** aller deutschen Frauen eine Körpergröße zwischen 156,4cm ($\mu - 2\sigma$) und 174,4cm ($\mu + 2\sigma$) haben
- **99,7%** aller deutschen Frauen eine Körpergröße zwischen 151,9cm ($\mu - 3\sigma$) und 178,9cm ($\mu + 3\sigma$) haben

Auf der Abbildung auf der rechten Seite ist die 68-95-99,7-Regel noch einmal grafisch veranschaulicht.

Tabelle der Populationen im Verhältnis zum Abstand der Standardabweichung zum Erwartungswert

Bereich	Population im Bereich	Erwartete Häufigkeit außerhalb des Bereichs	Ungefähres Auftreten, bei einem täglichen Ereignis
$\mu \pm 1\sigma$	0,682689492137086	1 aus 3	zwei Mal wöchentlich
$\mu \pm 1.5\sigma$	0,866385597462284	1 aus 7	wöchentlich
$\mu \pm 2\sigma$	0,954499736103642	1 aus 22	alle drei Wochen
$\mu \pm 2.5\sigma$	0,987580669348448	1 aus 81	vierteljährlich
$\mu \pm 3\sigma$	0,997300203936740	1 aus 370	jährlich
$\mu \pm 3.5\sigma$	0,999534741841929	1 aus 2.149	alle sechs Jahre
$\mu \pm 4\sigma$	0,999936657516334	1 aus 15.787	alle 43 Jahre (zweimal im Leben)
$\mu \pm 4.5\sigma$	0,999993204653751	1 aus 147.160	alle 403 Jahre
$\mu \pm 5\sigma$	0,999999426696856	1 aus 1.744.278	alle 4.776 Jahre (einmal in der bekannten Geschichtsschreibung)
$\mu \pm 5.5\sigma$	0,999999962020875	1 aus 26.330.254	alle 72.090 Jahre
$\mu \pm 6\sigma$	0,999999998026825	1 aus 506.797.346	alle 1,38 Millionen Jahre (Alter der Menschheit)
$\mu \pm 6.5\sigma$	0,999999999919680	1 aus 12.450.197.393	alle 34 Millionen Jahre
$\mu \pm 7\sigma$	0,99999999997440	1 aus 390.682.215.445	alle Milliarden Jahre
$\mu \pm x\sigma$	$\text{erf}\left(\frac{x}{\sqrt{2}}\right)$	1 aus $\frac{1}{1 - \text{erf}\left(\frac{x}{\sqrt{2}}\right)}$	alle $\frac{1}{1 - \text{erf}\left(\frac{x}{\sqrt{2}}\right)}$ Tage

Interaktive Normalverteilung

Durch Verändern des Erwartungswerts und der Standardabweichung verändert sich auch die Position bzw. die Form des Graphen der Normalverteilung. Dazu einfach die Werte durch beiden Regler unterhalb verändern.

Normalverteilung-Rechner

Der Rechner für die Normalverteilung berechnet den Wert der Verteilungsfunktion, der kumulierten Verteilungsfunktion sowie Quartile und kritische Werte.

$$P(X = k) = f(k; n, p) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

$$F(k; n, p) = P(X \leq k) = \sum_{i=0}^{\lfloor k \rfloor} \binom{n}{i} \cdot p^i \cdot (1 - p)^{n-i}$$

$$P(X \geq k) = \sum_{i=\lceil k \rceil}^n \binom{n}{i} \cdot p^i \cdot (1 - p)^{n-i}$$
