

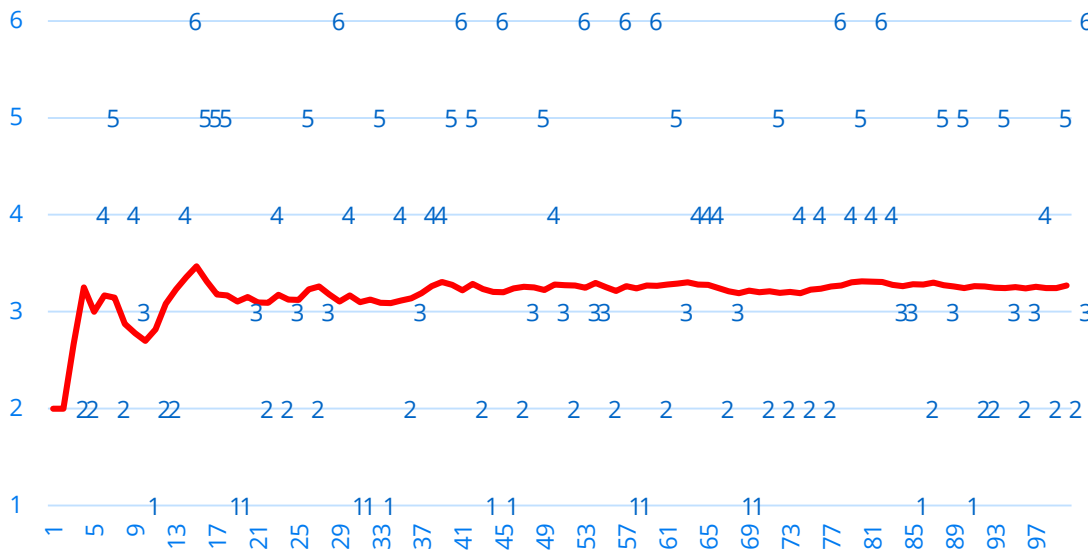
Zentraler Grenzwertsatz Normalverteilung

Was passiert, wenn wir unendliche Zufallsexperimente machen.

Erwartungswert und Zentrale Grenzwertsatz

Der Erwartungswert gibt an, welcher Wert zu erwarten ist, wenn das „Experiment“ unendlich oft wiederholt wird.

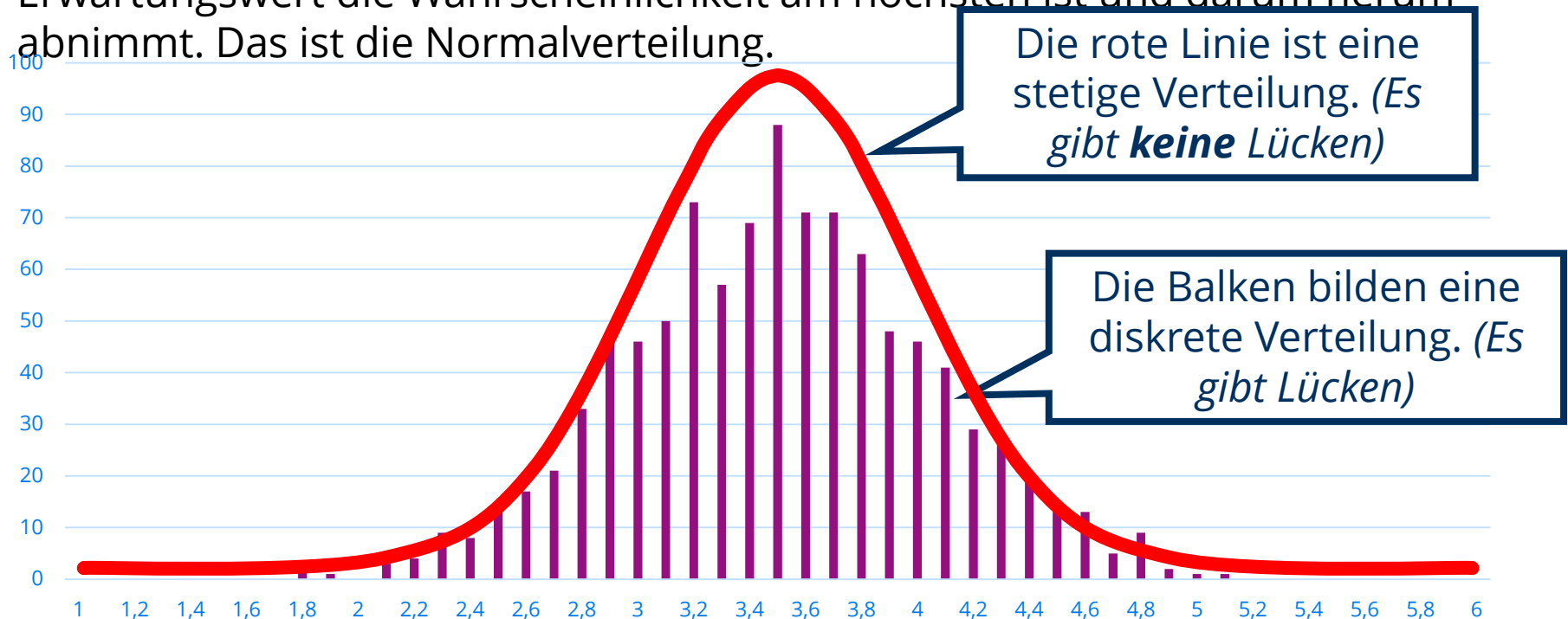
Die Grafik zeigt die Ergebnisse eines Würfelexperiments. Auf der X-Achse sind die 100 Ergebnisse der Reihe nach dargestellt. Auf der Y-Achse ist die Höhe der Augenzahl dargestellt. Die rote Linie zeigt den Durchschnitt, jeweils vom ersten Wurf an. Am Ende nähert sich dieser Durchschnitt dem Erwartungswert.



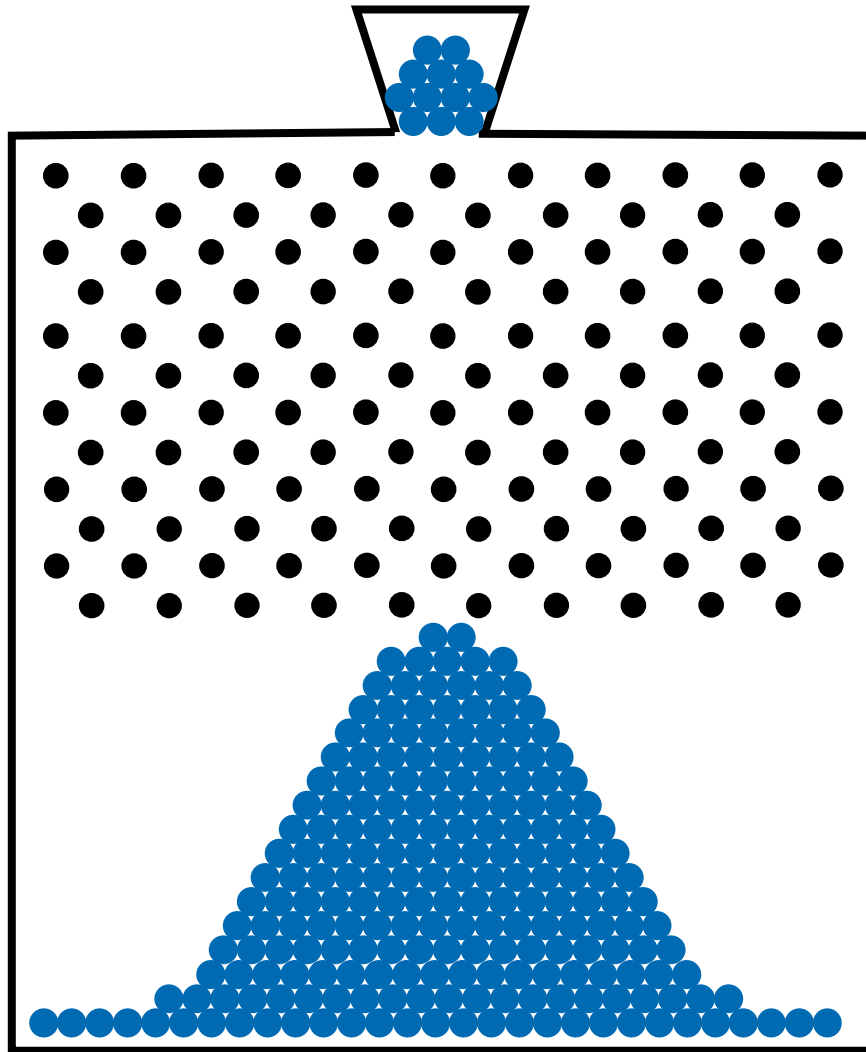
Führen Sie dieses Experiment selbst online durch:
<https://seeing-theory.brown.edu/basic-probability/index.html>
Scrollen Sie runter zu „Expectation“ und klicken Sie links „Roll 100 times. Wie viele Male müssen Sie auf den Button drücken, bis der Mittelwert sehr nah am Erwartungswert ist?“

Normalverteilung (diskrete und stetige Verteilung)

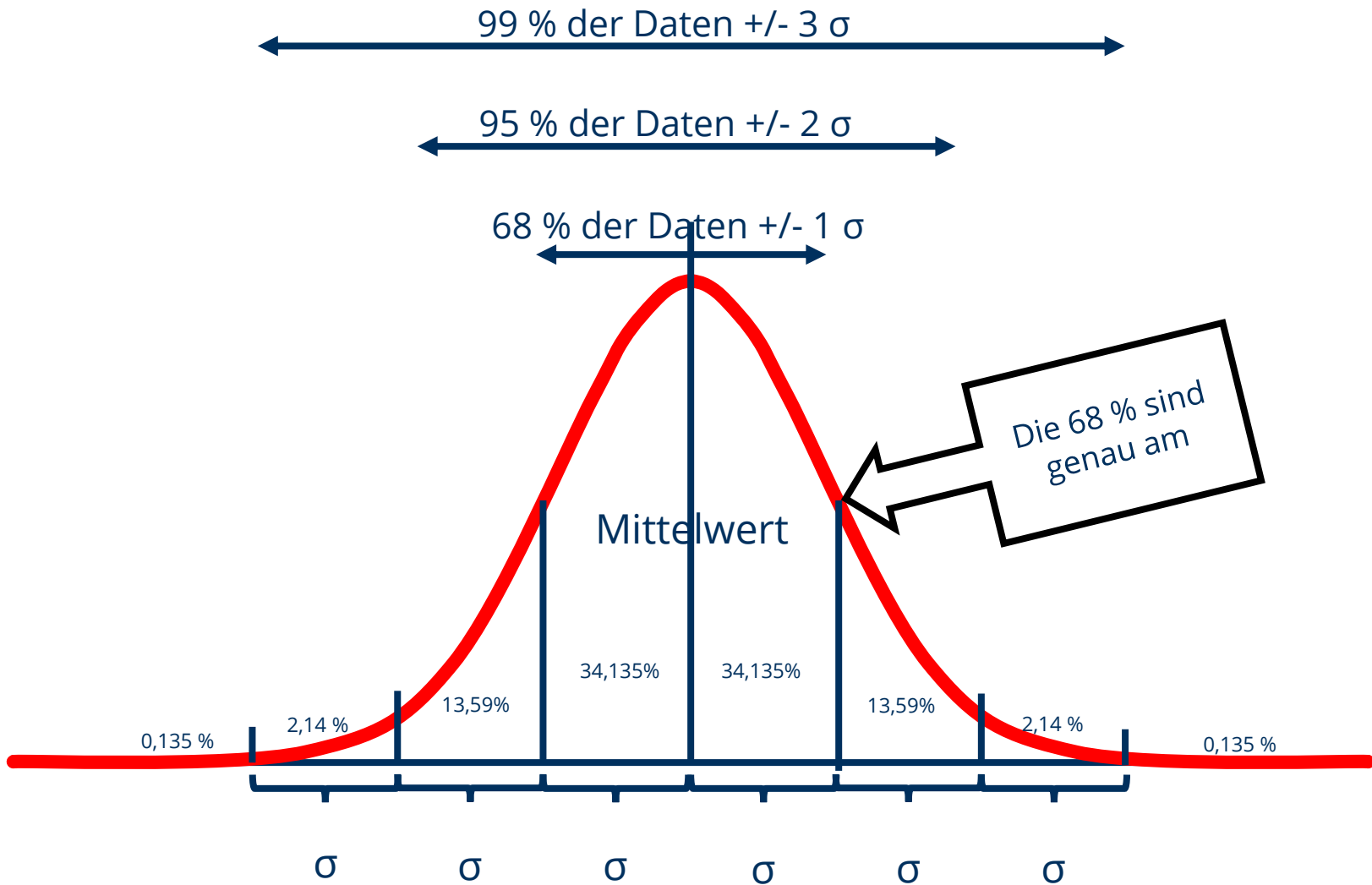
Wenn 1000 Stichproben mit 10 Würfeln gemacht werden. Jedes Mal der Mittelwert bilden. Dann kann man eine Verteilung beobachten, bei der um den Erwartungswert die Wahrscheinlichkeit am höchsten ist und darum herum abnimmt. Das ist die Normalverteilung.



Normalverteilung veranschaulichen: Galton-Brett



Normalverteilung und Standardverteilung



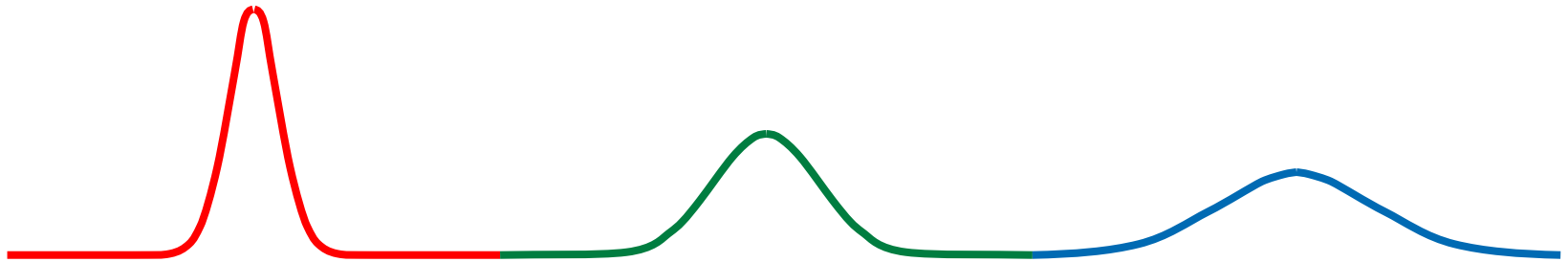
Übung zur Normalverteilung und Standardabweichung

In einem Datensatz (der eine glockenförmige Verteilung hat) lautet der Mittelwert 100 und die Standardabweichung 20.

- a. In welchem Wertebereich sollten 95 % der Daten liegen?
- b. Welcher Anteil an Werten wird in dem Bereich zwischen 80 und 120 liegen?
- c. Ab welchem Wert liegen 50 der Daten darüber?
- d. Wie viel Prozent der Werte liegen unter 40?

Formen der Normalverteilung

Eine Normalverteilung kann unterschiedliche Formen haben.

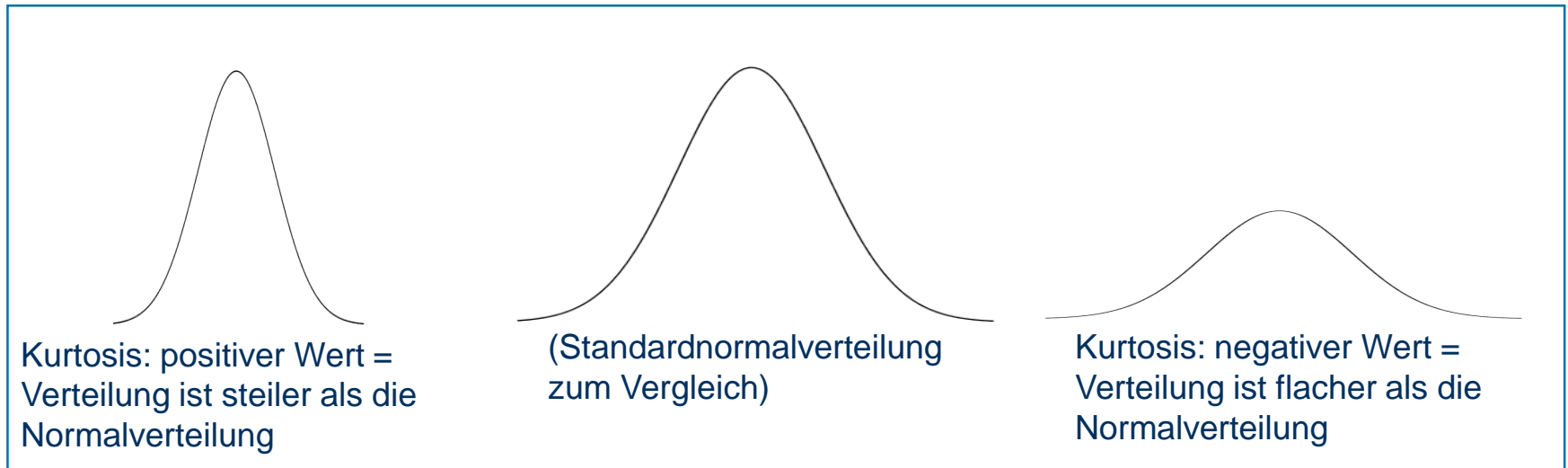


Trotzdem haben alle einige Eigenschaften

- ✓ Symmetrischer glockenförmiger Kurvenverlauf
- ✓ In der Mitte ein Hügel, am Ende sehr flach
- ✓ Mittelwert genau in Mitte
- ✓ Mittelwert = Median
- ✓ Standardabweichung beschreibt die Verteilung der Daten

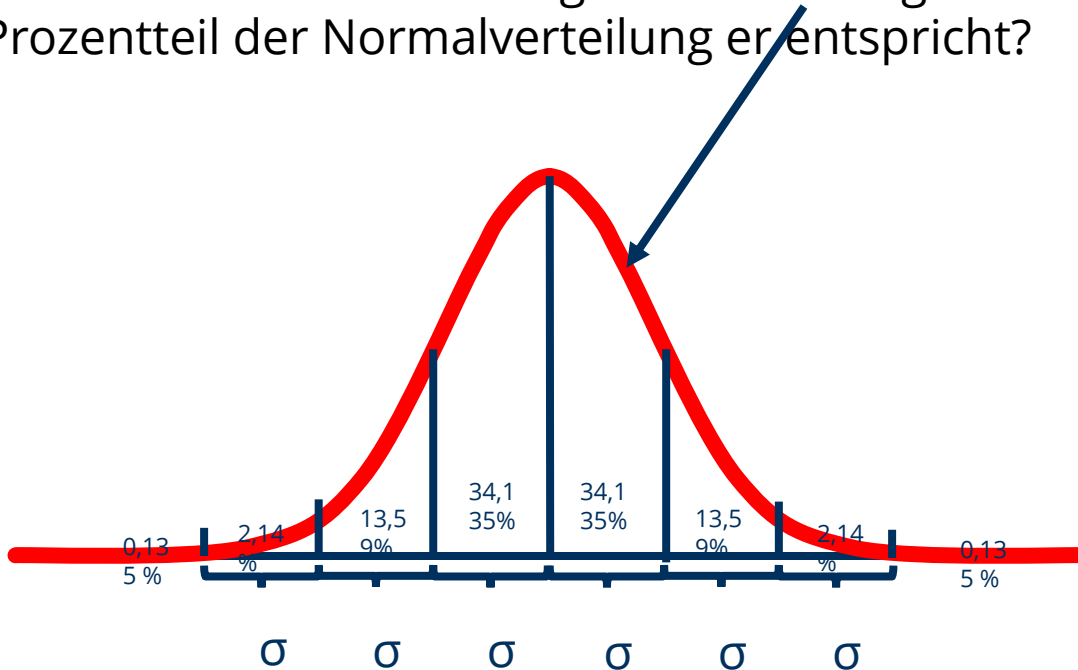
Verteilungsmaße: Wölbung (Kurtoisis oder Exzess)

Kurtosis: Beschreibt die Breite des Gipfels der Verteilung.

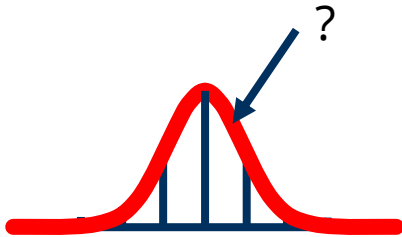


Beliebige Werte auf der Normalverteilungen und die entsprechenden Prozente ermitteln

Ich habe einen Wert der irgendwo hier liegt. Wie find ich raus, welchem Prozentteil der Normalverteilung er entspricht?



z-Standardisierung



Formel zur Standardisierung eines Wertes: $z = \frac{x - \mu}{\sigma}$

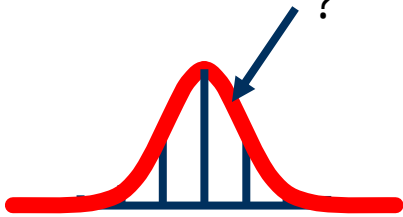
x = der beliebige Wert

μ = der Mittelwert

σ = Standardabweichung

Z-Wert in Tabelle ablesen siehe nächste Folie.

Standardwerte und Perzentile der Standardnormalverteilung (Z-Verteilung) **simple** Tabelle



Vom ermittelten z-Wert das Perzentil ablesen. (Bei negativen z-Werten den Prozentwert von 100 abziehen:
z.B. z-Wert = -0,7. $100 - 75,804 = 24,196 \%$)

z	Perzentil		z	Perzentil		z	Perzentil		z	Perzentil
0	50,000%		1	84,134%		2	97,725%		3	99,865%
0,1	53,983%		1,1	86,433%		2,1	98,214%		3,1	99,903%
0,2	57,926%		1,2	88,493%		2,2	98,610%		3,2	99,931%
0,3	61,791%		1,3	90,320%		2,3	98,928%		3,3	99,952%
0,4	65,542%		1,4	91,924%		2,4	99,180%		3,4	99,966%
0,5	69,146%		1,5	93,319%		2,5	99,379%		3,5	99,977%
0,6	72,575%		1,6	94,520%		2,6	99,534%		3,6	99,984%
0,7	75,804%		1,7	95,543%		2,7	99,653%		3,7	99,989%
0,8	78,814%		1,8	96,407%		2,8	99,744%		3,8	99,993%
0,9	81,594%		1,9	97,128%		2,9	99,813%		3,9	99,995%
									4	99,997%

Quelle: Wikipedia (2020) Standardnormalverteilungstabelle

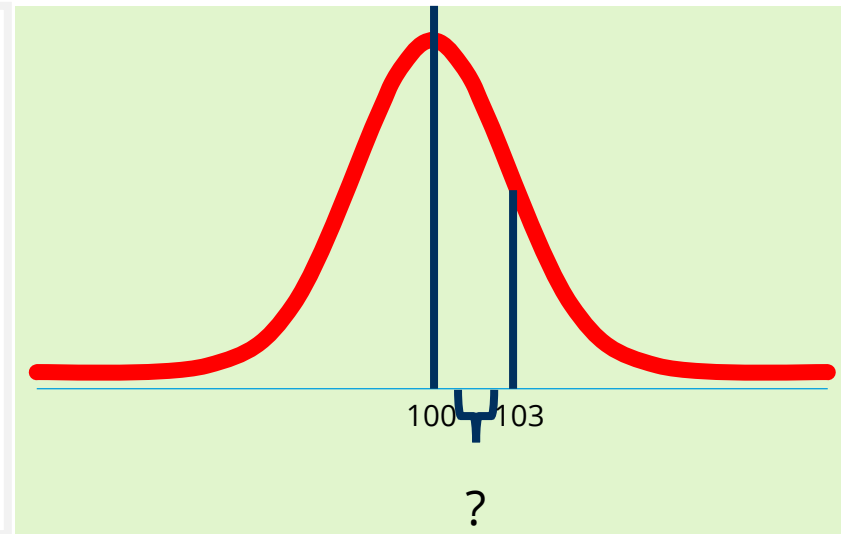
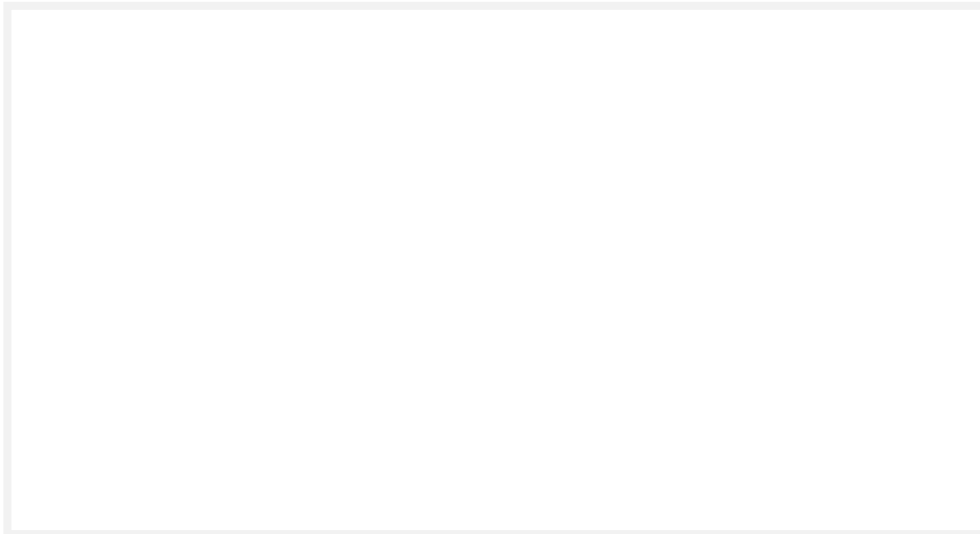
<https://de.wikipedia.org/wiki/Standardnormalverteilungstabelle> Datum: 08.04.2020

Übung: Rechnen mit Normalverteilung

Übungsbeispiel:

Gegeben: Mittelwert = 100, Standardabweichung = 3

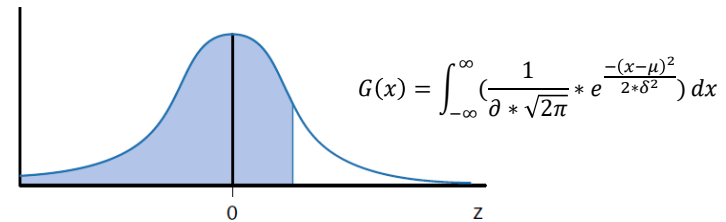
Gesucht: Wahrscheinlichkeit, dass x im Intervall von 101 bis 102 liegt.



Zur Ergänzung: <https://www.youtube.com/watch?v=khdJpwkmNf8>

Rechnen mit Normalverteilung: Verteilungsfunktion der Standardnormalverteilung in perfektionierter Tabelle

Quelle: Bortz, J.; Schuster, C. (2010).
Statistik für Human- und
Sozialwissenschaftler. Berlin, Heidelberg:
Springer.



Für den z-Wert von
2/3 (0,666) die
Wahrscheinlichkeit
ermitteln.

Zweite Dezimalstelle des z-Wertes

z-Wert*	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,60	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,70	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,80	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,90	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,00	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,10	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,20	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,30	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767

Von Perzentil über z-Wert zurück zum Originalwert

Mit der Normalverteilung kann man auch „rückwärts“ rechnen.

1. Ein Perzentilwert ist in der Aufgabenstellung gegeben + zwei weitere Werte (ein x-Wert, der Mittelwert oder die Standardabweichung)
2. Sie suchen den z-Wert in der Tabelle.
3. Mithilfe des z-Werts und den gegebenen Werten errechnen Sie den gesuchten (fehlenden) Wert. Z.B. wenn x gesucht ist: $x = z * \sigma + \mu$

Beispielaufgabe: Die Länge von Apfelbäumen ist normalverteilt mit dem Mittelwert von 4 Metern und einer Standardabweichung von 1 Meter. Welche Größe markiert die Grenze, ab der die 10 % der größten Bäume liegen?

$$x = 1,3 * 1 + 4 = \mathbf{5,3 \text{ Meter}}$$
 (1,3 kommen aus Tabelle)

Erstellen Sie eine Aufgabe in der der Perzentilwert vorgegeben ist und die Standardabweichung gesucht ist.

Normalverteilung: Interaktiv

Zur Wiederholung nutzen Sie das interaktive Tool auf dieser Seite:

<https://matheguru.com/stochastik/normalverteilung.html>

Was zeigt die Kumulative Verteilungsform (CDF) an?

Z-Werte in einer aktuellen Anwendung:

Auf der folgenden Internetseite finden Sie ebenfalls Z-Werte, die in der Corona-Krise häufiger den je angeschaut werden.

<https://www.euromomo.eu/graphs-and-maps/>

Finden Sie heraus, was die Z-Werte sagen!

Dieser Podcast (die ersten 7 Minuten) vom MDR wird Ihnen weiter helfen:

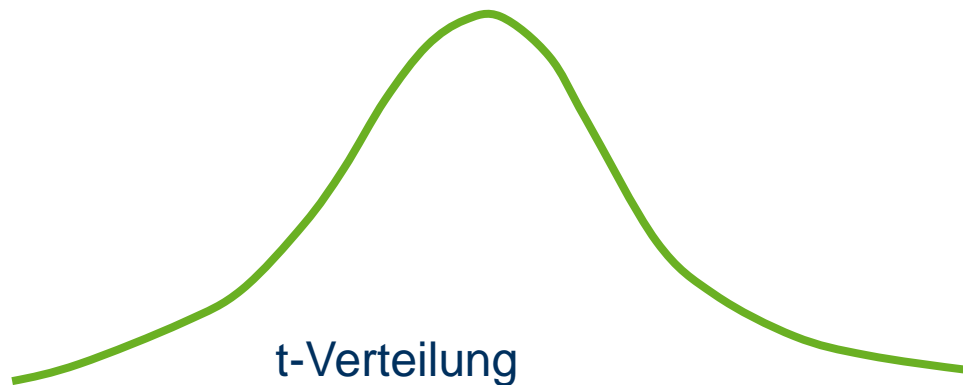
<https://www.mdr.de/nachrichten/podcast/kekule-corona/uebersterblichkeit-mortalitaet-remdesivir-nikotin-audio-100.html>

Kleine Stichproben $n < 30 \rightarrow$ t-Verteilung statt Normalverteilung

Die t-Verteilung hat eine etwas dickere Form als die Normalverteilung, was widerspiegelt, dass sie sich auf eine unsichere Datenbasis stützt.

Für die t-Werte gibt es eine eigene Tabelle, die von der Stichprobengröße abhängt.

Anzahl Freiheitsgrade n	75%	90%	95%	98%	98%	99%
1	1	3,078	6,314	12,706	31,821	63,657
2	0,816	1,886	2,92	4,303	6,965	9,925
3	0,765	1,638	2,353	3,182	4,541	5,841
4	0,741	1,533	2,132	2,776	3,747	4,604
5	0,727	1,476	2,015	2,571	3,365	4,032
6	0,718	1,44	1,943	2,447	3,143	3,707
7	0,711	1,415	1,895	2,365	2,998	3,499
8	0,706	1,397	1,86	2,306	2,896	3,355
9	0,703	1,383	1,833	2,262	2,821	3,25
10	0,7	1,372	1,812	2,228	2,764	3,169
11	0,697	1,363	1,796	2,201	2,718	3,106
12	0,695	1,356	1,782	2,179	2,681	3,055
13	0,694	1,35	1,771	2,16	2,65	3,012
14	0,692	1,345	1,761	2,145	2,624	2,977
15	0,691	1,341	1,753	2,131	2,602	2,947
16	0,69	1,337	1,746	2,12	2,583	2,921
17	0,689	1,333	1,74	2,11	2,567	2,898
18	0,688	1,33	1,734	2,101	2,552	2,878
19	0,688	1,328	1,729	2,093	2,539	2,861
20	0,687	1,325	1,725	2,086	2,528	2,845
21	0,686	1,323	1,721	2,08	2,518	2,831
22	0,686	1,321	1,717	2,074	2,508	2,819
23	0,685	1,319	1,714	2,069	2,5	2,807
24	0,685	1,318	1,711	2,064	2,492	2,797
25	0,684	1,316	1,708	2,06	2,485	2,787
26	0,684	1,315	1,706	2,056	2,479	2,779
27	0,684	1,314	1,703	2,052	2,473	2,771
28	0,683	1,313	1,701	2,048	2,467	2,763
29	0,683	1,311	1,699	2,045	2,462	2,756
30	0,683	1,31	1,697	2,042	2,457	2,75
z-Werte	0,674	1,282	1,645	1,96	2,326	2,576



Die t-Verteilung wird uns auch beim t-Test wieder begegnen.

Konfidenzintervall

Warum Konfidenzintervalle

Aussage: Im Durchschnitt machen die Deutschen 32,0 Überstunden im Jahr.

Problem: das trifft nicht für jeden zu, besser wäre die Angabe eines Intervalls. Es ist unwahrscheinlich, dass ein Mittelwert* aus einer Stichprobe mit dem „wahren“ Mittelwert in der Grundgesamtheit übereinstimmt.

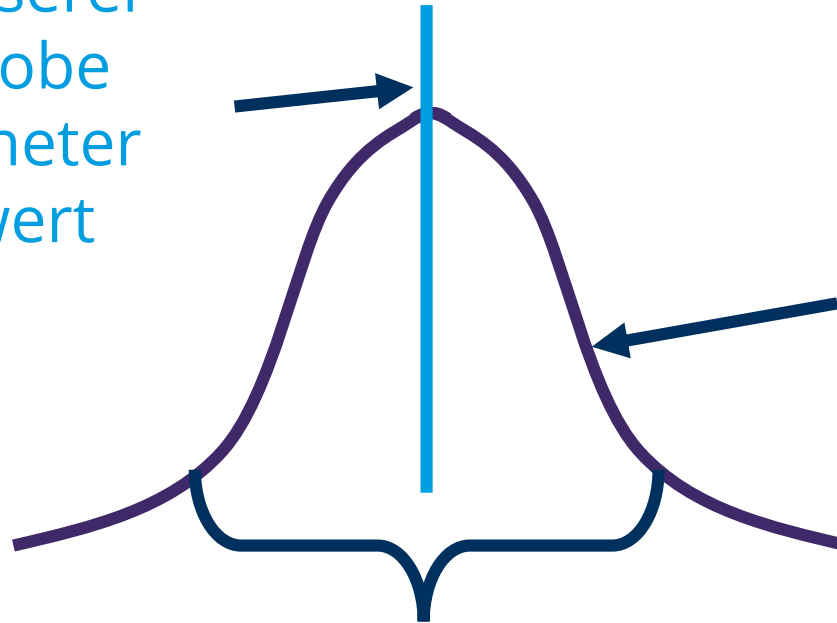
Lösung: Deshalb gibt man einen Bereich an, in dem sich der „wahre“ Mittelwert“ befindet. In diesem Bereich ist der Mittelwert dann zu einem bestimmten Prozentsatz zu finden – meist 95 Prozent.

Der Bereich heißt **Konfidenzintervall**

*anstatt des Mittelwertes können auch für andere Werte Konfidenzintervalle gebildet werden (Anteilswerte, Regressionskoeffizienten)

Konfidenzintervall: Darstellung

Aus unserer
Stichprobe
errechneter
Mittelwert



Die
Wahrscheinlichkeit
für den wahren
Mittelwert. Achtung
nicht die Daten!

Im Konfidenzintervall befindet sich der „wahre“ Mittelwert mit einer Wahrscheinlichkeit von 95 %.

Berechnung des Konfidenzintervalls

$$\text{Konfidenzintervall für den Stichprobenmittelwert} = \pm Z * \frac{\sigma}{\sqrt{n}}$$

Z ist wieder der Z-Wert aus der Normalverteilungstabelle für das gewünschte Wahrscheinlichkeitsintervall. Meist:

$$90 \% = 1,64$$

$$95 \% = \mathbf{1,96}$$

$$99 \% = 2,58$$

Übungsaufgabe Konfidenzintervall

Übungsaufgabe: Gesucht ist das **95 % Konfidenzintervall** für eine Stichprobe mit einer **Standardabweichung von 30** und dem Mittelwert von 100, bei einer **Stichprobe von 900**.

$$\text{Fehlergrenze} = \pm 1,96 * \frac{30}{\sqrt{900}} = 1,96$$



Nun ziehen Sie die Fehlergrenze einmal vom Mittelwert (100) ab und Sie erhalten

→ Die untere Grenze des Konfidenzintervalls = 98,04

Um die obere Grenze zu erhalten addieren Sie das Ergebnis einmal auf den Mittelwert hinzu:

→ Obere Grenze des Konfidenzintervalls = 101,96

Interpretation Konfidenzintervall

A. Achtung man interpretiert das Konfidenzintervall nicht so:

Mit 95 % Sicherheit liegt der wahre Mittelwert in dem Konfidenzintervall.

Sondern:

- In 95 % aller vergleichbar ermittelten Stichproben liegt der Mittelwert zu 95 % in diesem Intervall.
- Oder: *Wir wissen/glauben/vermuten ...* mit 95 % Sicherheit, das der wahre Mittelwert in den Grenzen des Konfidenzintervalls liegt.

B. Beachten Sie, den Unterschied zwischen Fehlergrenze und obere/untere Schranke/Grenze des Konfidenzintervalls. (siehe vorherige Folie)

C. Die Einheiten des Konfidenzintervalls sind gleich die Einheiten des Stichprobenmittelwerts. Z.B. beim Mittelwert der Körpergröße z.B. in Zentimetern, kann das Konfidenzintervall z.B. 10 Zentimeter betragen.

Konfidenzintervall Übung

Berechnen Sie das 95 % Konfidenzintervall für Rennstrecken von 100 Sportlern in einer Woche:

Mittelwert: 200 km

Standardabweichung: 40

Erläutern Sie was das Konfidenzintervall bedeutet.

Übung Konfidenzintervall abhängig von Verteilung

Nutzen Sie interaktive Grafik: <https://seeing-theory.brown.edu/frequentist-inference/index.html#section2>

Scrollen Sie zu „Confidence Intervall“

Suchen Sie als Verteilungsformen die Normalverteilung aus.

Drücken Sie nun auf „Start Sampling“.

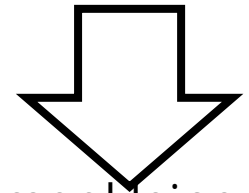
Beobachten Sie nun etwa eine halbe Minute wie das Programm zufällige Stichproben ($n = 5$) aus der Verteilung zieht und jeweils das Konfidenzintervall um den Mittelwert herum berechnet.

Wählen Sie im Anschluss eine neue Verteilung (z.B. uniform) und starten das Experiment erneut.

Beschreiben Sie wie sich die Konfidenzintervalle zwischen beiden Verteilungsformen unterscheiden.

Eigenschaften der Fehlergrenze

$$\text{Fehlergrenze für Mittelwert} = \pm Z * \frac{\sigma}{\sqrt{n}}$$



Wenn das n – also die Stichprobe – größer wird, wird der Fehler immer kleiner.

Daraus lässt sich ableiten, dass größere Stichproben zu verlässiger Aussagen über den gefundenen Mittelwert zulassen.

Was passiert mit der Fehlergrenze, wenn das Konfidenzniveau (Z) größer wird?

Was passiert mit der Fehlergrenze, wenn die Standardabweichung größer wird?

Kann es eine Fehlergrenze geben in der zu 100 % der Wert liegt?

Konfidenzintervall Fragen

	a	b
Sie haben für eine Stichprobenmittelwert zwei Konfidenzintervalle errechnet ein a.) 95 % und b.) ein 99 %. Welches ist breiter?		
Sind breitere Konfidenzintervalle a.) erstrebenswerter oder b.) eher nicht erstrebenswerter als schmalere Konfidenzintervalle?		
Mit dem Konfidenzintervall von 95 % wissen wir a.) zu 100 % das der Mittelwert der Stichprobe in diesem Intervall liegt b.) zu 95 % das der Mittelwert der Stichprobe in diesem Intervall liegt.		
Bei einer Stichprobe mit dem Mittelwert 50 und einer Fehlergrenze von 2 hat das Konfidenzintervall a.) eine Breite von 2 oder b.) eine Breite von 4 Einheiten.		
Um die Fehlergrenze bei einem 95 % Konfidenzintervall zu verringern kann man a.) die Stichprobe erhöhen oder b.) die Standardabweichung erhöhen.		

Konfidenzintervall: Konfidenz in Excel

=KONFIDENZ.NORM(Alpha;Standardabweichung;Umfang)

Der Standardwert für Alpha ist 0,05.
Am Ende der Rechnung erhalten wir
so das 95-Prozent-
Konfidenzintervall.

Die Standardabweichung
errechnet sich aus den
Rohdaten mit der Funktion
=STABW.S(Bereich)

Der Umfang ist einfach
die Stichprobengröße.

Werte			
1	Mittelwert	4,5	
2	Standardabweichung:	2,449	=STABW.S(A2:A9)
3	Stichprobengröße:	8	
4	Konfidenz für 95 %:	1,697	=KONFIDENZ.NORM(0,05; D3;8)
5			
6			
7			
8			

Konfidenzintervall: Konfidenzintervall in Excel

Da das Konfidenzintervall ein Intervall ist muss die „Konfidenz“ noch einmal auf den Mittelwert addiert und einmal subtrahiert werden.

Beispiel:

Konfidenz = 1,687 (siehe vorherige Folie)

Mittelwert = 4,5

Obere Grenze: = Mittelwert + Konfidenz
= 4,5 + 1,687 = 6,197

Untere Grenze: = Mittelwert – Konfidenz
= 4,5 – 1,687 = 2,803