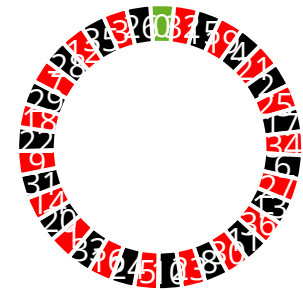
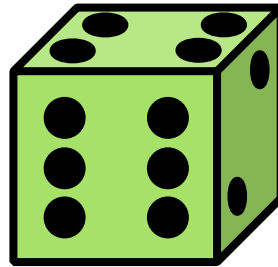


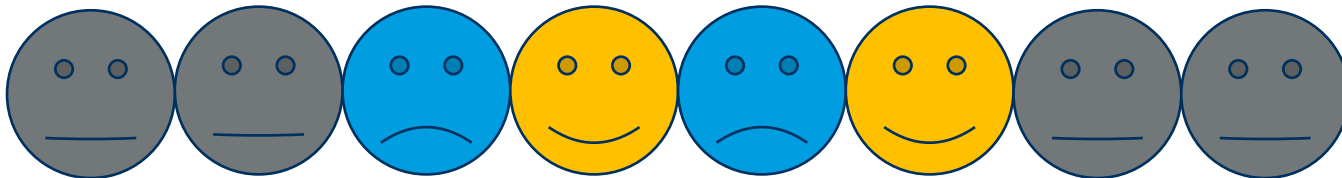
Stichproben

Von Wahrscheinlichkeit zur Statistik

Bei Münzen, Würfeln und Roulettespielen sind die Wahrscheinlichkeiten für das Auftreten eines konkreten Ereignisses **bekannt**.



Bei anderen Beispielen (aus Sozialwissenschaft) sind die Wahrscheinlichkeiten **nicht eindeutig bekannt**, dann reden wir von Statistik mit Stichproben.



Stichproben, Grundgesamtheit

Das ist unsere Grundgesamtheit (Population), alle Personen eine Gruppe



Wir ziehen zufällig 10 Personen, das ist unsere Stichprobe:



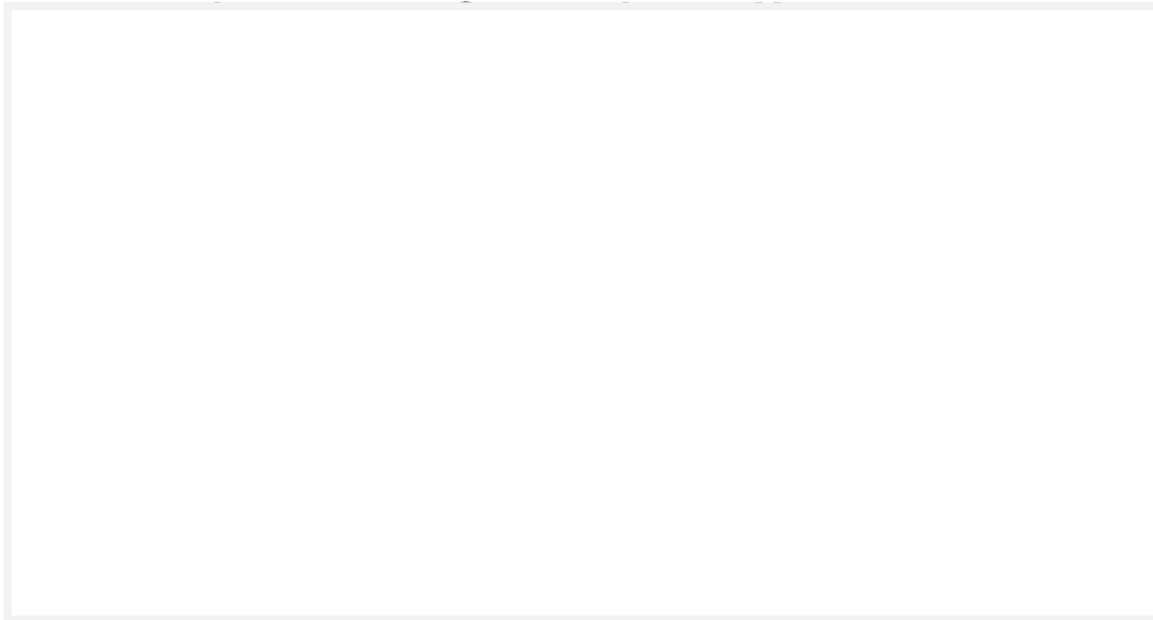
Wir analysieren diese Stichprobe und machen Rückschlüsse auf die Grundgesamtheit.

Wahrscheinlichkeit für gelb ist $\frac{\text{Anzahl Gelber}}{\text{Anzahl der Elemente in der Stichprobe}} = \frac{8}{10} = 80 \%$

Warum stimmen die 80 % aus der Stichprobe mit der Grundgesamtheit nicht über ein? Wie sicher können wir die Rückschlüsse und Interpretationen aus der Stichprobe auf die Grundgesamtheit ziehen?

Einführung in die Stichprobentheorie

Warum werden Stichproben gezogen?



Übung zu Grundgesamtheit und Stichprobe

1. Eine Politikerin möchte wissen, wie viel ältere Menschen wöchentlich Sport machen. Was ist ihre Grundgesamtheit?
2. Eine Wissenschaftlerin möchte den Zusammenhang untersuchen zwischen der Motivation und Arbeitsleistung. Sie macht Experimente mit Psychologiestudenten. Was ist ihre Grundgesamtheit und ist die Stichprobe geeignet? Welchen Einfluss kann die Stichprobe auf das Ergebnis haben?
3. Auf einer Internetseite werden Sie gefragt, wenn Sie wählen würden, wenn nächsten Sonntag Wahl wäre? Wer soll die Grundgesamtheit sein? Aus welcher Grundgesamtheit wird die Stichprobe tatsächlich gewählt?
4. Eine Initiative möchte das Gesundheitsverhalten von sehr armen Menschen analysieren. Sie ziehen dazu eine Stichprobe aus dem Melderegister. Ist es eine Zufallsauswahl? Wer ist die Grundgesamtheit, wenn erreichen Sie.

Arten von Stichproben

Einfache Zufallsstichprobe:


Geschichtete Auswahl:

Klumpenstichprobe:

Mehrstufig Auswahl:

Bewusste Auswahl:

Schneeballsystem:

- 
- All diesen Verfahren haben eine Regel
 - Sie sind nicht willkürlich
 - Die Grundgesamtheit ist immer angebar

Übung Stichproben

1. Sie messen den Benzinpreis in Ihrer Stadt an 25 Tankstellen und kommen auf einen Mittelwert von 1,20 Euro. Sie schließen nun, dass in Ihrem Bundesland der Mittelwert bei 1,20 Euro liegt. Was stimmt hierbei nicht?
2. Der Verein für eine „sportliche Stadt“ hat unter seinen Mitgliedern eine Umfrage zum Thema Ausbau der Sportstätten gemacht. Das Ergebnis: 90 % der Menschen sprechen sich für mehr Investition in die Sportstätten aus. Was stimmt hierbei nicht?
3. Eine Kosmetikartikelhersteller hat anhand einer Stichprobe ($n = 7$) herausgefunden, dass sein Produkt zu einem jüngeren Äußeren führt und das dies auch zwei Stunden nach auftragen des Produkts noch so ist. Was stimmt hier nicht?
4. Ein Marktforschungsinstitut hat auf einer Einkaufsstraße 1000 Leute nach Ostern befragt und herausgefunden, dass kaum einer der Befragten in den nächsten 5 Tagen Schokolade essen möchte. Was stimmt hier nicht?
5. Ein Trainerin testet, ob Hallentraining oder Outdoortraining besser ist. Dazu lässt sie die Teilnehmenden freiwillig entscheiden, in welche Gruppe sie möchten. Was stimmt hier nicht?
6. Eine Firma wirbt mit einer Lieferzeit von 3 Tagen. Der Verbraucherschutz macht nach Beschwerden eine Stichprobe und kommt auf eine durchschnittliche Lieferzeit von 5 Tagen. Hat die Firma gelogen?

Wieso entstehen zufällige Ergebnisse?

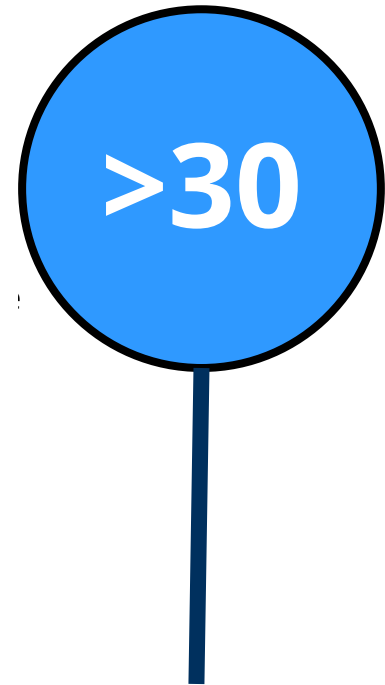
... weil keine Totalerhebung der Grundgesamtheit gemacht wird, sondern Stichproben gezogen werden.

Beispiel: In unserer Stichprobe erscheinen nur die roten Zahlen

Grundgesamtheit FH	Grundgesamtheit BA	Stichprobe FH	Stichprobe BA
500 €	500 €	500 €	500 €
500 €	500 €	500 €	500 €
500 €	500 €	500 €	
500 €	500 €	500 €	
1.000 €	1.000 €	1.000 €	1.000 €
1.000 €	1.000 €	1.000 €	1.000 €
1.000 €	1.000 €	1.000 €	1.000 €
1.000 €	1.000 €	1.000 €	
1.500 €	1.500 €	1.500 €	1.500 €
1.500 €	1.500 €	1.500 €	1.500 €
1.500 €	1.500 €		1.500 €
1.500 €	1.500 €		
2.000 €	2.000 €	2.000 €	2.000 €
2.000 €	2.000 €		2.000 €
2.000 €	2.000 €		2.000 €

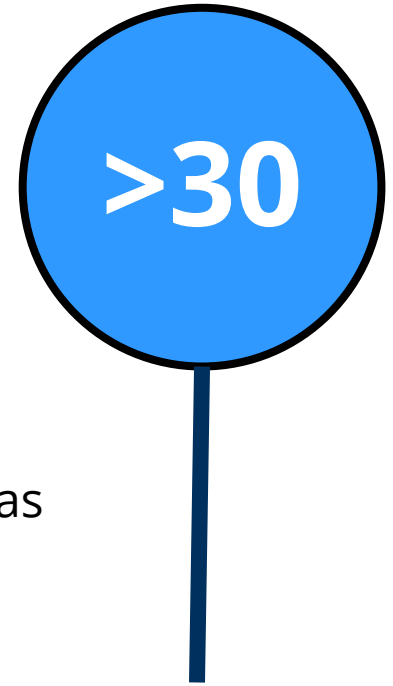
	FH	BA
Unsere Stichprobe		
In Wirklichkeit		

Aus der Praxis: Stichprobengröße berechnen in der Praxis und warum wir das nicht lernen



2. Grund warum eine Stichprobe von $n > 30$?

In einer vorhergehenden Übung haben wir 30 Stichproben gezogen mit einer jeweiligen Größe von 30? Warum?



Wir werden es (so gut wie) nie mit Variablen in den Sozialwissenschaften haben, die wirklich normalverteilt sind. Das heißt wir könnten viele Berechnungen (die wir später in der Veranstaltung machen wollen) nicht rechnen, weil dafür die Voraussetzung ist, dass die Daten normalverteilt sind. Das wär ungünstig. Daher gibt es eine Ausnahme (die empirische mathematisch auch begründet ist): Wenn Ihre Stichprobe mindestens 30 Fälle hat, dann kann man mindestens annehmen das ein aus so einer Stichprobe gewonnener Mittelwert normalverteilt ist und man schön weiter rechnen kann.

(Wenn es um die Berechnung von Anteilswerten (p) geht gibt es eine andere Formel: $p \cdot n > 5$.)

Übung: Unnötige Prozente

In einer Analyse eines Fragebogens steht die Aussage: „Der Anteil der Männer (40 %) war in der Stichprobe geringer als der von Frauen (60 %).“ Wie entsteht die Aussage bei einer Gesamtzahl von 5 Befragten? Warum ist es nicht sinnvoll, diese Aussage zu machen?

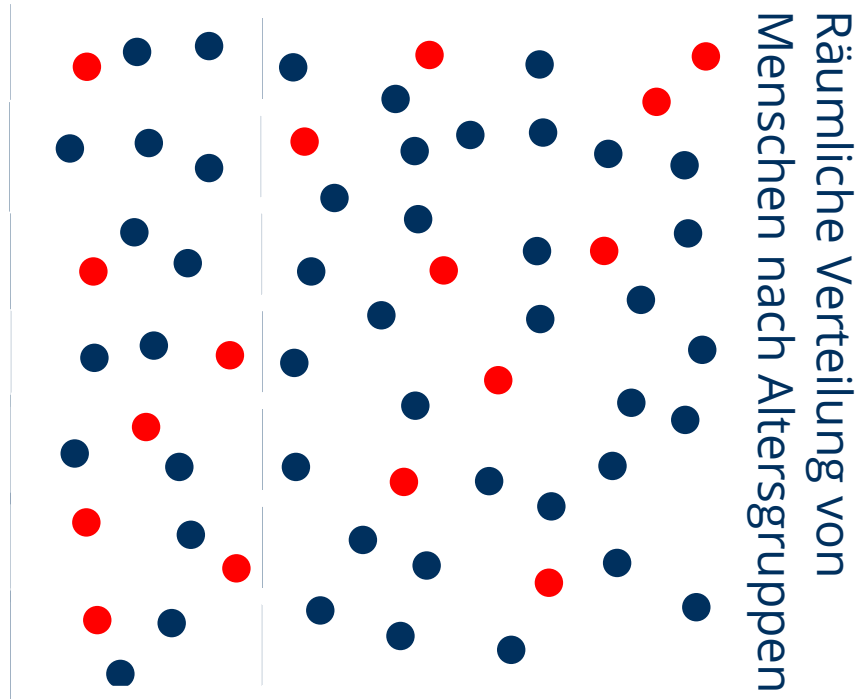
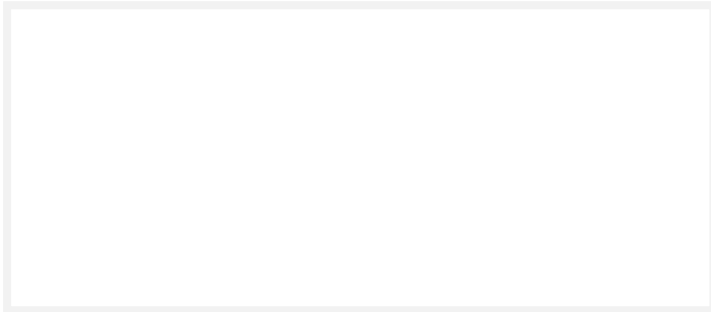
Rohdaten:

Nummer	Geschlecht
1	
2	
3	
4	
5	

Einen Mythos lüften

Stichprobengröße: Ein kleines Dorf mit sehr alten Leuten

- Es gibt kleine Dörfer in dem sehr viele Menschen sehr alt sind.
- Liegt das am gesunden Lebensstil oder auch an der Statistik?



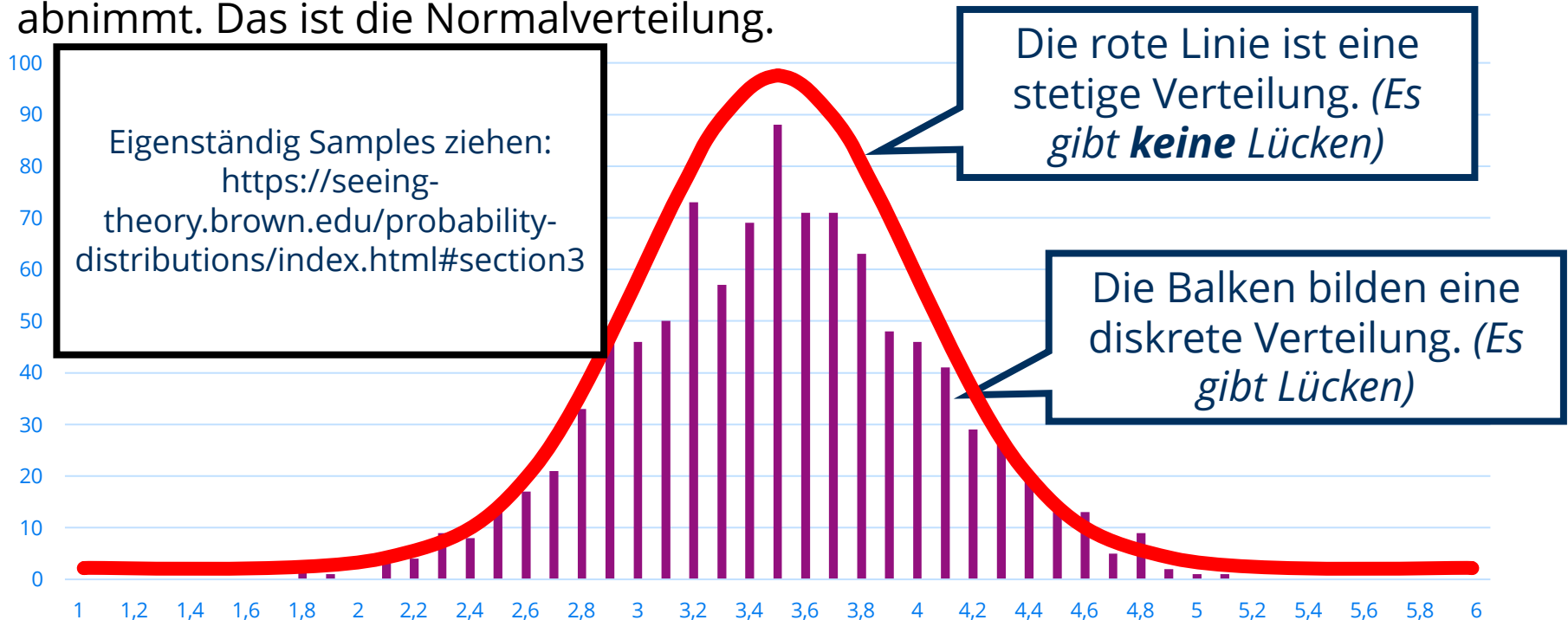
- Extrem alte Menschen
- Restlichen Menschen

Zentraler Grenzwertsatz Normalverteilung

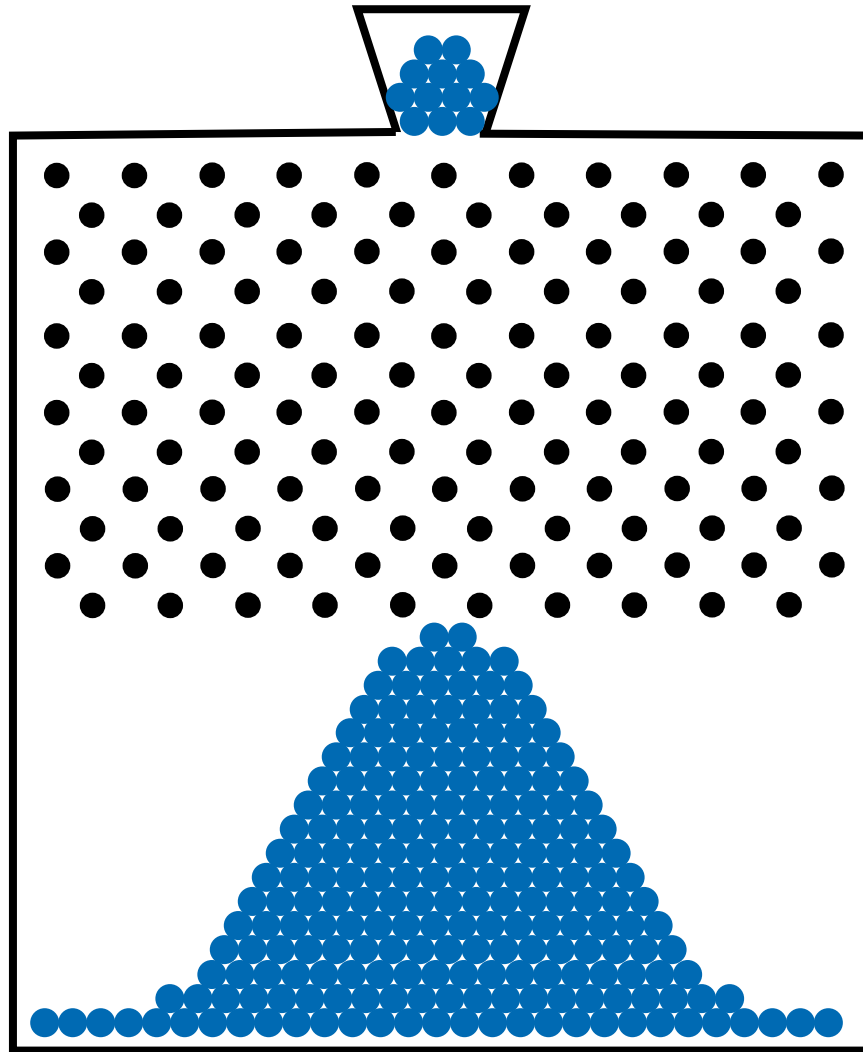
Was passiert, wenn wir unendliche Zufallsexperimente machen.

Normalverteilung (diskrete und stetige Verteilung)

Wenn Sie 1000 Stichproben mit 10 Würfeln machen und jedes Mal den 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.



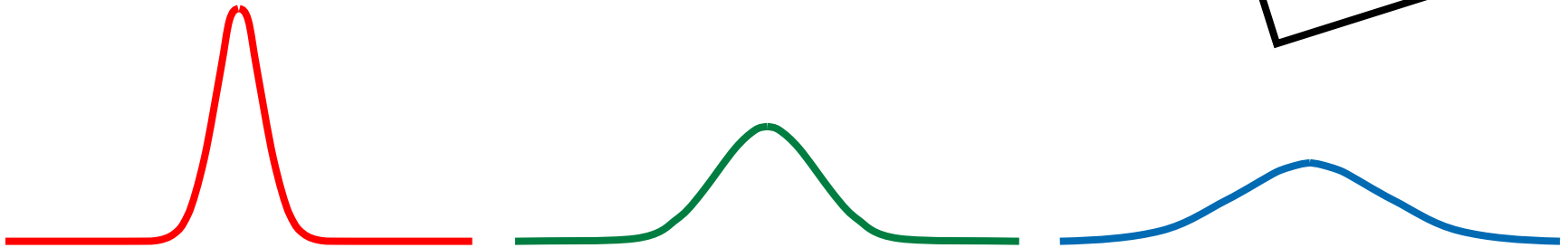
Normalverteilung veranschaulichen: Galton-Brett



Eigenschaften der Normalverteilung

Normalverteilte Daten lassen sich mit Schiefe und Wölbung beschreiben. Siehe unter Streuungsmaße.

Eine Normalverteilung kann unterschiedliche Formen haben.

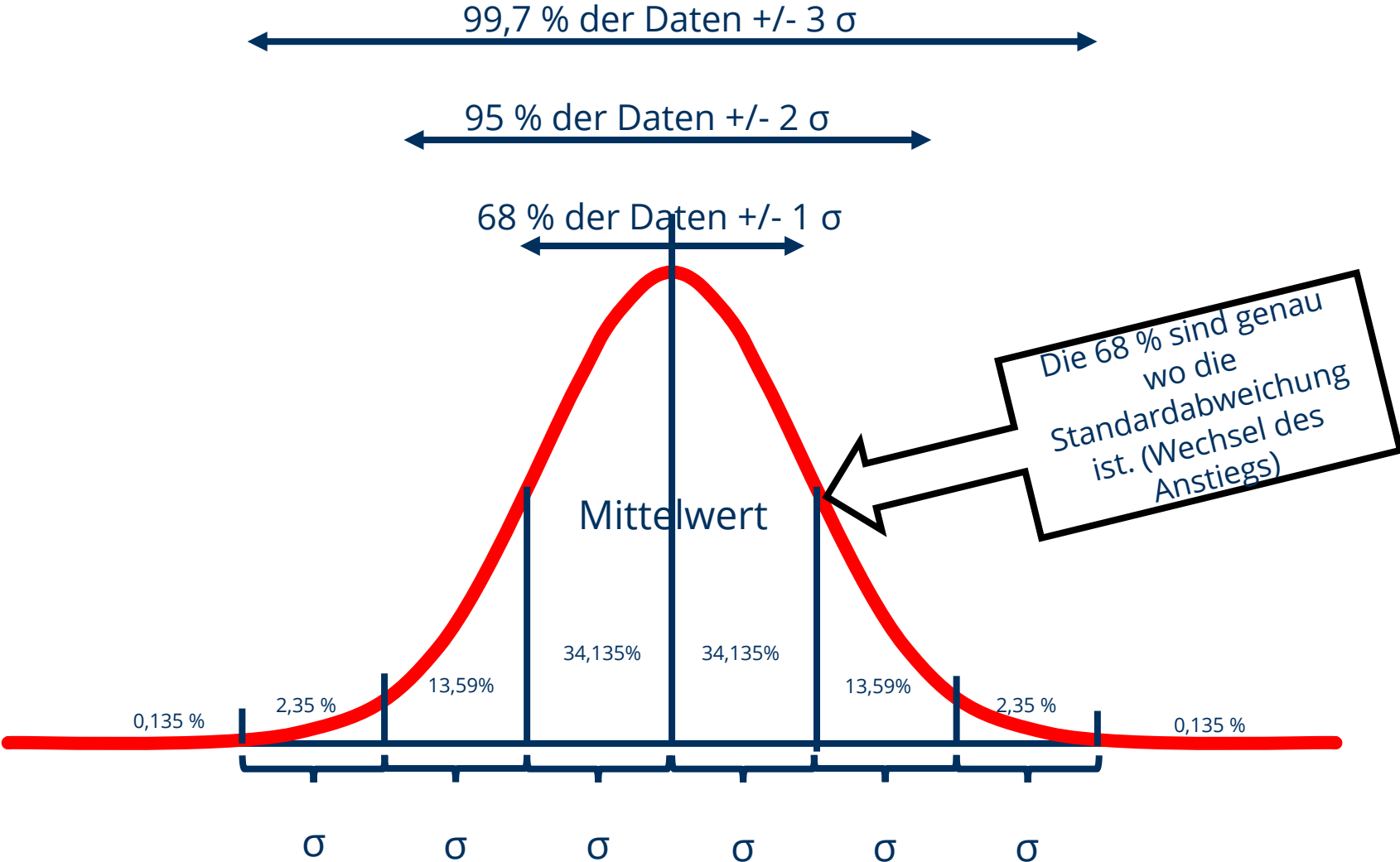


Trotzdem haben alle bestimmte Eigenschaften



Rechnen mit der Normalverteilung

Normalverteilung und Standardnormalverteilung



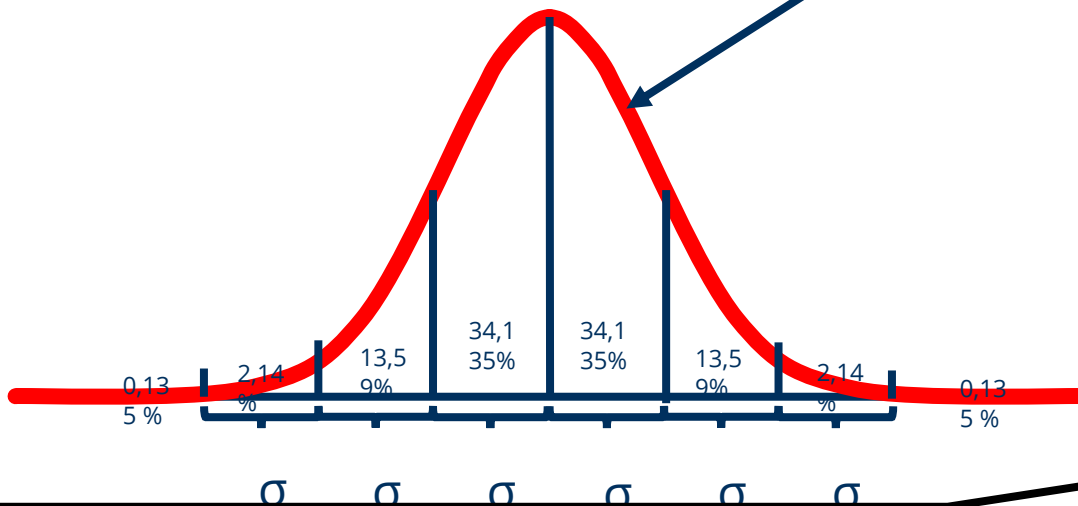
Übung zur Normalverteilung und Standardabweichung

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 dem x-Wert von 40?

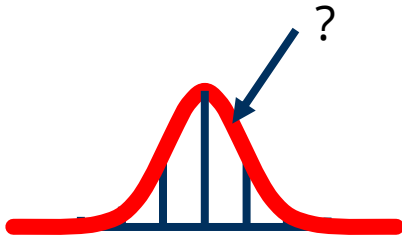
Beliebige Werte auf der Normalverteilungen und die entsprechenden Prozente ermitteln

Ich will nun aber den Prozentwert zu diesem x-Wert wissen.



- z-Wert für den x-Wert errechnen mit einer Formel (siehe nächste Seite)
- Prozentwert für den z-Wert in einer Tabelle ablesen

Die z-Standardisierung

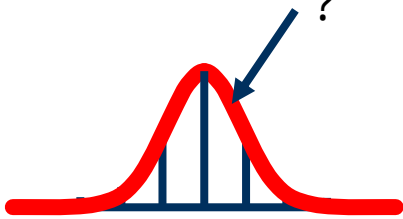


Formel zur Standardisierung eines Wertes

$$z = \frac{x - \bar{x}}{s}$$

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

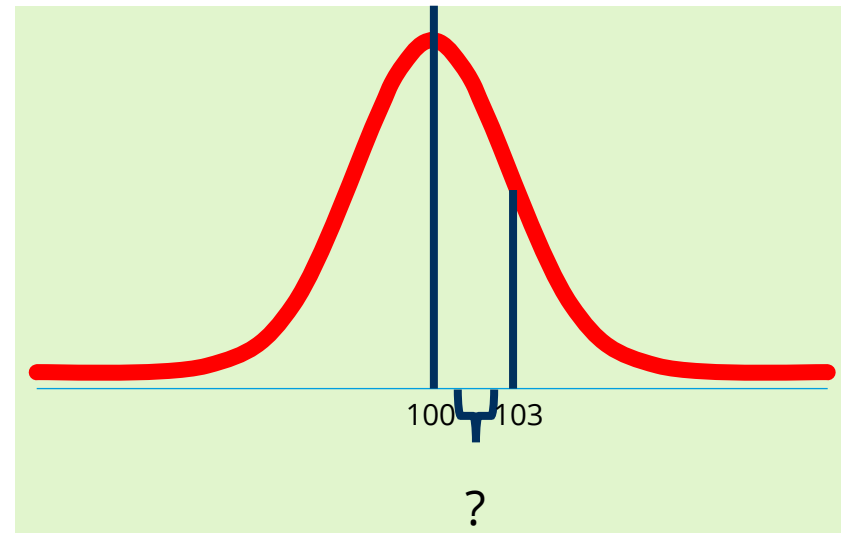
Gesucht: Wahrscheinlichkeit, dass ein beliebiges x im Intervall von 101 bis 102 liegt.

$z_1 = \frac{102-100}{3} = \frac{2}{3} = 0,666$; → zugehörigen z – Wert in Tabelle suchen = 72,58 %

$z_2 = \frac{101-100}{3} = \frac{1}{3} = 0,333$; → zugehörigen z – Wert in Tabelle suchen = 61,76 %

= 72,58 – 61,76 %

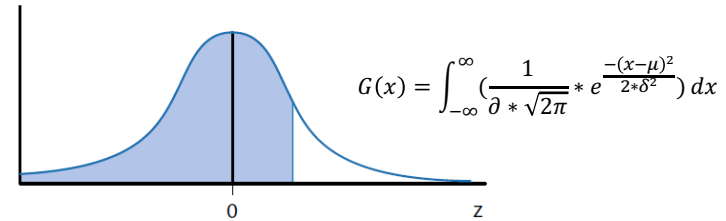
= 10,82 %



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 Prozentwert wird vorgegeben und zwei weitere Werte (z.B. Standardabweichung und Mittelwert, oder auch ein konkreter x-Wert)
2. Sie suchen in der Tabelle den z-Wert für den Prozentwert.
3. Mithilfe aller gegebenen Werte und des ermittelten z-Werts ermitteln Sie den gesuchten Wert (z.B. x), dazu stellen Sie die Formel um: $x = z \cdot \sigma + \mu$

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?

In der Tabelle finden wir für 10 % den z-Wert = 1,3

$$X = 1,3 \cdot 1 + 4 = 5,3 \text{ Meter}$$

Übung: Erstellen Sie eine Aufgabe in der die Standardabweichung gesucht wird und der Prozentwert und Mittelwert gegeben sind.

Normalverteilung: Interaktiv

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/> (unter Z-scores by country)

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>

Zur Wiederholung nutzen Sie das interaktive Tool auf dieser Seite:

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

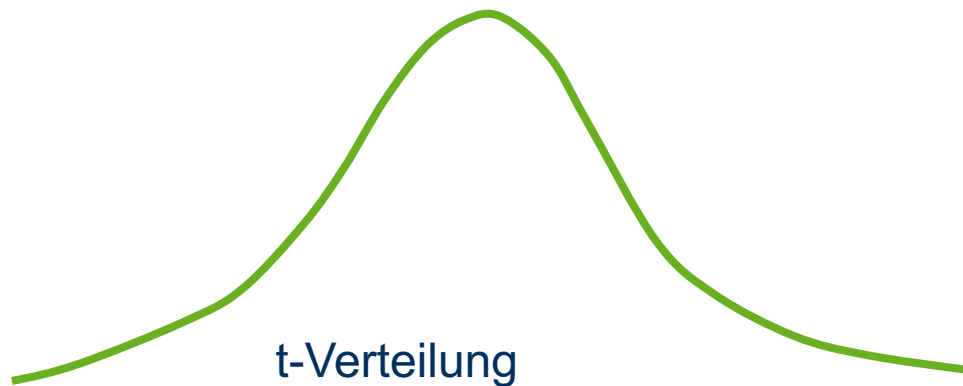
Was zeigt die Kumulative Verteilungsform (CDF) an?

Beispiel für eine andere Verteilung/kleine Stichproben: Wenn $n < 30$ nimmt man die 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.

Wiederholungsübung: Zentrale Grenzwertsatz, Normalverteilung, Stichproben

Machen Sie eine richtige oder falsche Aussage zu einem der obigen Begriffe.

<https://padlet.com/MariaNeubauerDresden/3iuga2iive9qxjba>

Wenn ausreichend viele Bewertungen eingegangen sind, dann bewerten sie die Aussagen mit Daumen nach oben oder unten.