STICHPROBE UND GRUNDGESAMTHEIT

Grundgesamtheit (N)

Gesamtmenge aller Einheiten (= Fälle), die das untersuchungsrelevante Merkmal bzw. untersuchungsrelevante Merkmalskombinationen aufweisen.

Stichprobe (n)

Teilmenge aller Einheiten (= Fälle) der Grundgesamtheit, die das untersuchungsrelevante Merkmal bzw. untersuchungsrelevante Merkmalskombinationen aufweisen.

➤ Stichprobentheorie, Stichprobenziehungsverfahren

Inferenz

Schluss von Stichprobendaten auf „wahre“ Parameter der Grundgesamtheit -> Parameterschätzung

Signifikanz

Sicherheit bzw. Zufälligkeit / Nicht-Zufälligkeit der Schätzung

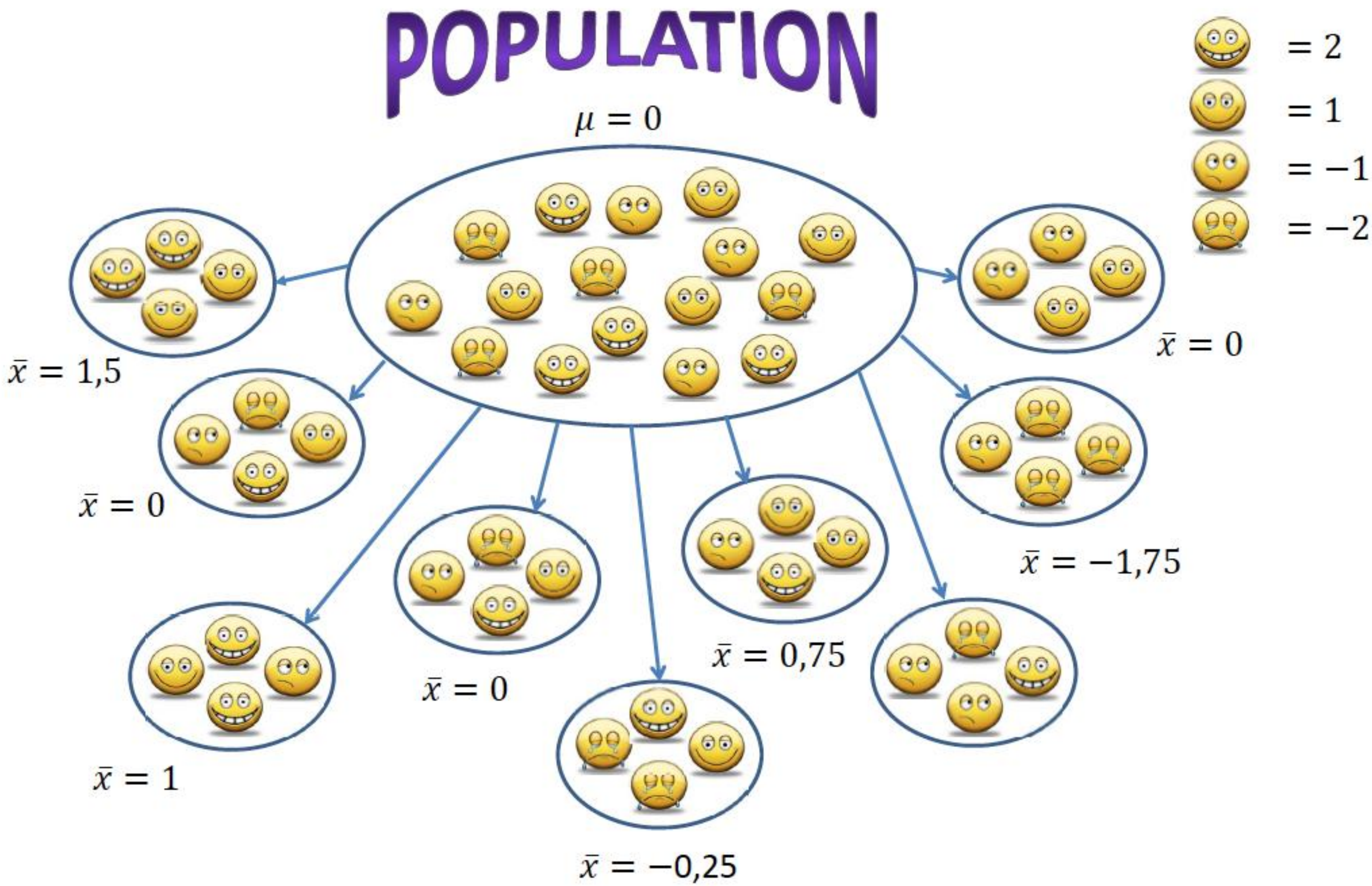
Ziel

Schätzung unbekannter Parameter (Verteilungen, Kennwerte) in der Grundgesamtheit auf der Basis bekannter Statistiken (Verteilungen, Kennwerte) von Stichprobenerhebungen.

	Kennwert der Stichprobe	Parameter der Grundgesamtheit
Mittelwert	$\overline{x}$	μ
rel. Häufigkeit	p	π
Varianz	s^2	σ^2
Standardabweichung	s	σ
Korrelation	r	ρ

Schluss von Stichproben auf Grundgesamtheit: Warum und wie darf man das?

Beispiel: Mittelwerte in Stichproben



Stichprobenkennwerte

Stichprobenverteilung

„Die statistischen Maßzahlen, die man für einzelne Stichproben berechnet, heißen **Stichprobenkennwerte** (Mittelwert, Varianz, Prozentanteil). Die Verteilung der Stichprobenkennwerte aus Stichproben heißt (...) Stichprobenverteilung."

(Diaz-Bone 2006: 142)

Die **Stichprobenkennwerte** sind Schätzer für die entsprechenden Kennwerte der Grundgesamtheit.

Schätzung: Annahme

Der Stichprobenkennwert entspricht dem Kennwert in der Grundgesamtheit mit einem **Fehler** von e.

Beispiele:

Schätzung des arithmetischen Mittels μ, i Stichproben

$$\bar{x}_1 + e_1 = \mu$$

$$\bar{x}_2 + e_2 = \mu$$

(...)

$$\boxed{\bar{x}_i + e_i = \mu}$$

Schätzung des Prozentanteils p, i Stichproben

$$p_1 + e_1 = \pi$$

$$p_2 + e_2 = \pi$$

(...)

$$\boxed{p_i + e_i = \pi}$$

Zufallsvariablen und Zufallsexperiment

Zufallszahlen

Stichprobenkennwerte können im wahrscheinlichkeitstheoretischen Sinne als **Zufallszahlen** betrachtet werden.

Das heißt: Erkenntnisse aus Stichproben über die Grundgesamtheit können mittels wahrscheinlichkeitstheoretischer Modelle gewonnen werden.

Im statistischen Zusammenhang spricht man von **Zufallsvariablen**.

Definitionen

Eine Variable X, die jedem möglichen Ergebnis x_i eines **Zufallsexperiments** eine reelle Zahl zuordnet, wird als **Zufallsvariable** bezeichnet.

Ein **Zufallsexperiment** muss (a) beliebig wiederholbar, (b) nach bestimmten, gleichen Regeln durchgeführt, und (c) im Ergebnis unsicher bzw. „zufällig" sein.

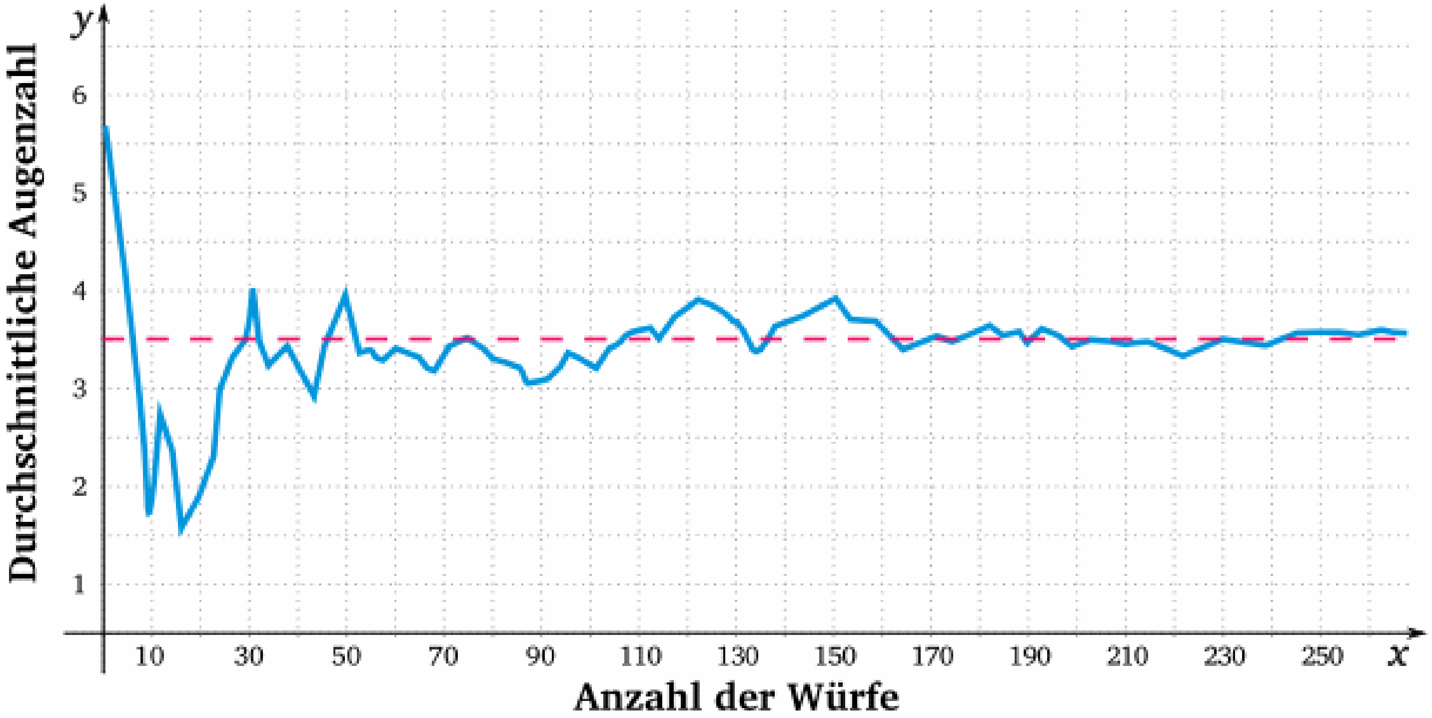
Stichprobentheorie

Die **Wahrscheinlichkeitsverteilung** einer Zufallsvariable X ist die Zuordnung aller Wahrscheinlichkeiten zu allen durch X festgelegten Ereignissen.

Der **Erwartungswert** einer Zufallsvariable ist der Wert, den sie im Mittel vieler Wiederholungen annimmt („durchschnittlich zu erwartender Wert").

Beispiel

Würfeln. Notiert werden die geworfenen Zahlen und Anzahl der Würfe, berechnet wird jeweils die durchschnittliche Augenzahl. Bei gleicher Wahrscheinlichkeit für jede Zahl beträgt der **Erwartungswert** für die durchschnittliche Augenzahl 1+2+3+4+5+6 / 6 = 3,5. (rote Linie)



Quelle:
https://matheguru.com/stochastik/gesetz-der-grossen-zahlen.html

Anderes Beispiel

Münzwurf. Die theoretische Wahrscheinlichkeit (= der Erwartungswert) für Kopf bzw. Zahl beträgt ½ = 0,5.

Anzahl Würfe	davon Kopf		Verhältnis		absoluter Abstand	relativer Abstand
	theoretisch	beobachtet	theoretisch	beobachtet		
100	50	48	0.500	0.480	2	0.020
1000	500	491	0.500	0.491	9	0.009
10000	5000	4970	0.500	0.497	30	0.003

Wird ein Zufallsexperiment immer wieder unter denselben Bedingungen durchgeführt, so nähert sich das Ergebnis immer weiter der Wahrscheinlichkeit des Zufallsexperiments an.

oder auch:

Eine tatsächliche Verteilung nähert sich mit steigender Fallzahl immer mehr der theoretisch erwarteten Verteilung an.

Das Gesetz der Großen Zahlen

Wahrscheinlichkeitsverteilungen

Wahrscheinlichkeitstheorie

Grundbegriffe: Zufallsvariable, Zufallsexperiment, Erwartungswert...

Wahrscheinlichkeitsverteilungen

Diskrete Zufallsvariablen

Eine Variable X, zwischen deren **abzählbaren Ausprägungen** keine Zwischenwerte existieren.

Bei diskreten Zufallsvariablen kann man nach der Wahrscheinlichkeit des Auftretens bestimmter Werte fragen.

Stetige Zufallsvariablen

Eine Variable X, bei der zwischen zwei Ausprägungen **unendlich viele Ausprägungen** existieren.

Bei stetigen Zufallsvariablen kann man nach dem Auftreten von Werten innerhalb eines bestimmten Intervalls fragen.

Wahrscheinlichkeitsfunktion und Dichtefunktion

Definitionen

Eine Wahrscheinlichkeitsfunktion gibt für jedes Zufallsereignis, also jeden x_i -Wert einer **diskreten** Zufallsvariable X an, mit welcher Wahrscheinlichkeit er auftritt.

Die Wahrscheinlichkeitsfunktion f ordnet jedem x_i -Wert eine Wahrscheinlichkeit P zu. Die Summe der Einzelwahrscheinlichkeiten ist immer gleich 1.

$$f(x_i) = p_i \qquad \sum_i f(x_i) = \sum_i p_i = 1$$

Bei einer **stetigen** Zufallsvariable X geht die Wahrscheinlichkeit jedes einzelnen möglichen Zufallsereignisses gegen Null. Daher werden Wahrscheinlichkeiten für Intervalle bestimmt.

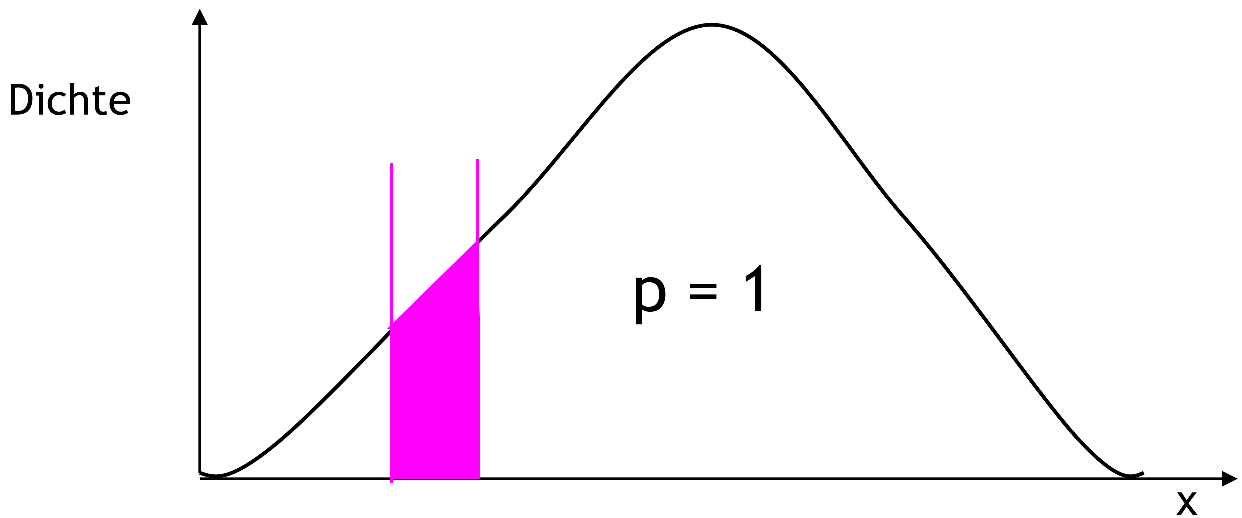
Die Dichtefunktion für alle denkbaren Werte (von -∞ bis +∞) ist 1.

$$p\ (a \leq x \leq b) = \int_a^b f(x)\, dx$$

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

Graphik Dichtekurve

Die Gesamtfläche unter der Dichtekurve wird gleich 1 gesetzt, der Flächenanteil des entsprechenden Intervalls $\boxed{\Delta x}$ entspricht dann der gesuchten Wahrscheinlichkeit



Stetige Wahrscheinlichkeitsverteilung

Wahrscheinlichkeitsverteilungen

Wahrscheinlichkeitstheorie

Grundbegriffe: Zufallsvariable,
Zufallsexperiment, Erwartungswert...

Wahrscheinlichkeitsverteilungen

Diskrete Zufallsvariablen	Stetige Zufallsvariablen
↓	↓
Binomialverteilung	Normalverteilung

Binomialverteilung 1

Definition

Die Binomialverteilung ist eine diskrete Wahrscheinlichkeitsverteilung, in deren Rahmen Ereignisse in zwei Alternativen auftreten. Die alternativen Ereignisse können gleich oder ungleich wahrscheinlich sein.

Parameter

Anzahl der Fälle (n)
Wahrscheinlichkeit (p), Gegenwahrscheinlichkeit (q bzw. 1-p)

Eigenschaften

symmetrisch, falls $p = q = 0,5$

asymmetrisch, falls $p \neq q$

Mit wachsender Fallzahl (n) nähern sich die Verteilungen den erwarteten Werten an.

→ **Gesetz der Großen Zahl**

Definition

Anzahl der Gesamtversuche (Fälle)

Wahrscheinlichkeit für das Ereignis

Die Wahrscheinlichkeit für Ereignis x bei einem Zufallsexperiment

$$\boxed{P(x)} = \binom{n}{x} \boxed{p}^{x} * (\boxed{1-p})^{n-x}$$

Gegenwahrscheinlichkeit (1-p)

Anzahl der gewünschten Ereignisse

$$\frac{n!}{(n-x)! * x!}$$

Binomialkoeffizient

(alternative Schreibweise mit x = K)

Anwendungsbeispiel 1

Wie hoch ist die Wahrscheinlichkeit in einer Familie mit 3 Kindern, dass genau ein Kind ein Mädchen ist?

3 Kinder: n = 3

2 potentielle Ereignisse (Ausprägungen): weiblich, männlich

Wahrscheinlichkeit p = 0,5

ein erwünschtes Ereignis (Mädchen): x = 1

$$P(x) = \binom{n}{x} p^x * (1-p)^{n-x}$$

$$= \binom{3}{1} 0{,}5^1 * 0{,}5^{3-1}$$

$$= \frac{3!}{(3-1)!*1!} * 0{,}5^1 * 0{,}5^2$$

$$= \frac{3!}{(2)!*1!} * 0{,}5 * 0{,}25$$

$$= \frac{3 * 2 * 1}{2 *1*1} * 0{,}125 = 0{,}375 = \frac{3}{8}$$

Anwendungsbeispiel 2

Wie hoch ist die Wahrscheinlichkeit für zwei defekte Produkte, wenn für die Qualitätsprüfung aus der Fertigung 5 Produkte entnommen werden, und die Wahrscheinlichkeit für fehlerhafte Produkte 0,05 beträgt?

$$P(x) = \binom{n}{x} p^x * (1-p)^{n-x}$$

$$= \binom{5}{2} 0{,}05^2 * 0{,}95^{5-2}$$

$$= \frac{5!}{(5-2)! * 2!} * 0{,}05^2 * 0{,}95^3$$

(...)

$$= 0{,}0124$$

5 entnommene Produkte: n = 5

Wahrscheinlichkeit Fehler p = 0,05

2 erwünschter Ereignisse (Fehlerprodukte): x = 2

Quelle: https://www.frustfrei-lernen.de/mathematik/binomialverteilung.html

Definition

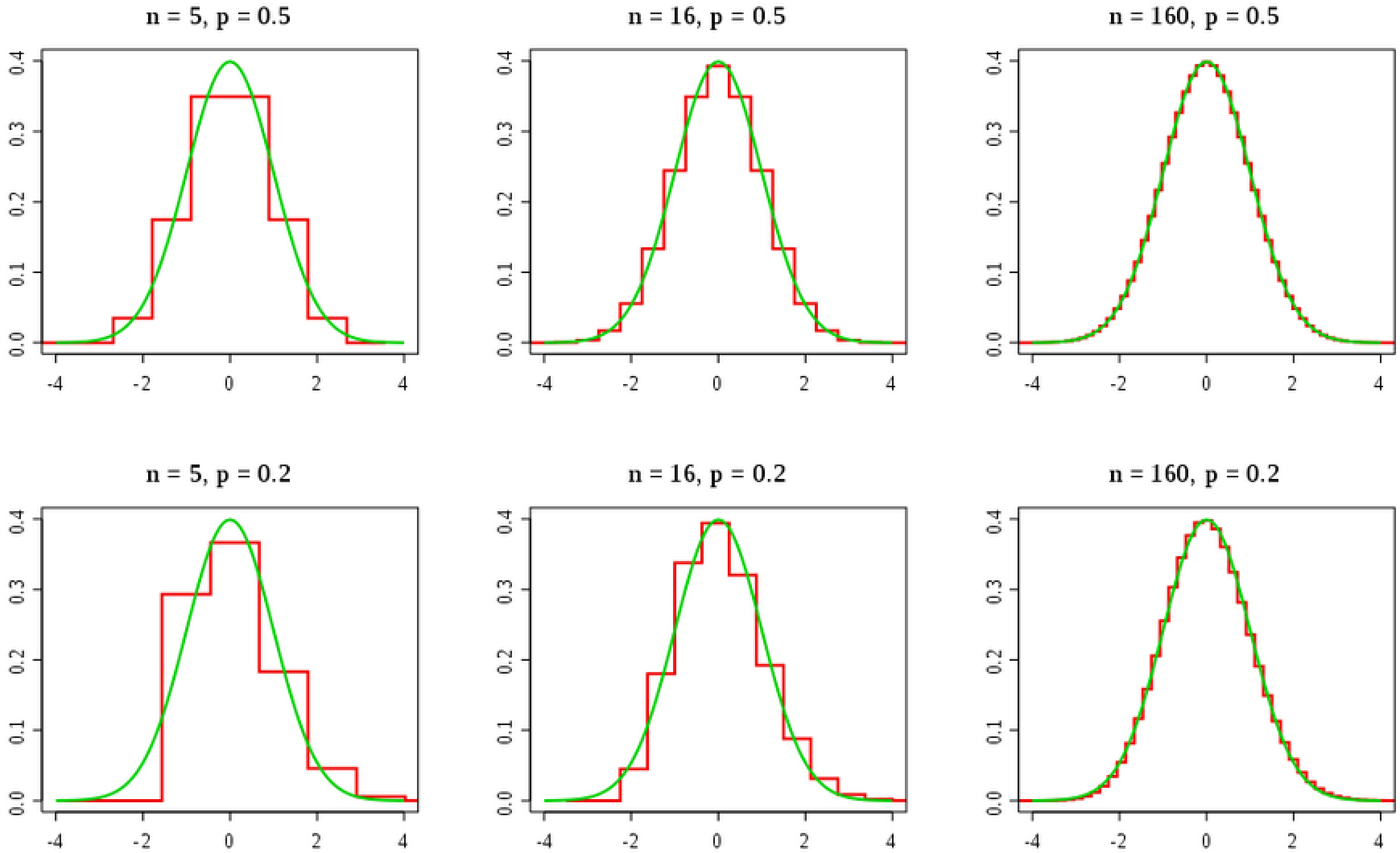
Die *Mittelwertverteilung in Stichproben* ist asymptotisch *normalverteilt*, und zwar unabhängig von der Form der zugrundeliegenden Verteilung der Daten, wenn die Daten unabhängig und identisch verteilt sind.

oder einfacher:

Die Verteilung von *Mittelwerten aus Stichproben* des Umfangs n, die sämtlich derselben Grundgesamtheit entnommen wurden, geht mit wachsendem Stichprobenumfang in eine *Normalverteilung* über.

Zentraler Grenzwertsatz

Zentraler Grenzwertsatz



Quelle: Wikipedia/Sinner1

Wahrscheinlichkeitsverteilungen

Wahrscheinlichkeitstheorie

Grundbegriffe: Zufallsvariable,
Zufallsexperiment, Erwartungswert…

Wahrscheinlichkeitsverteilungen

Diskrete Zufallsvariablen

Binomialverteilung

Stetige Zufallsvariablen

Normalverteilung

Normalverteilung

Definition

Wichtigste Verteilungsform stetiger Zufallsvariablen, bei denen zwischen zwei Ausprägungen **unendlich viele Ausprägungen** existieren.

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

$-\infty < x < +\infty, \sigma > 0$

Parameter

Erwartungswert (μ)

Varianz (σ^2) bzw. Standardabweichung (σ)

$N \; (x, \mu, \sigma)$

Eigenschaften

glockenförmige Form

symmetrisch, Symmetrieachse $x = \mu$ = Maximum

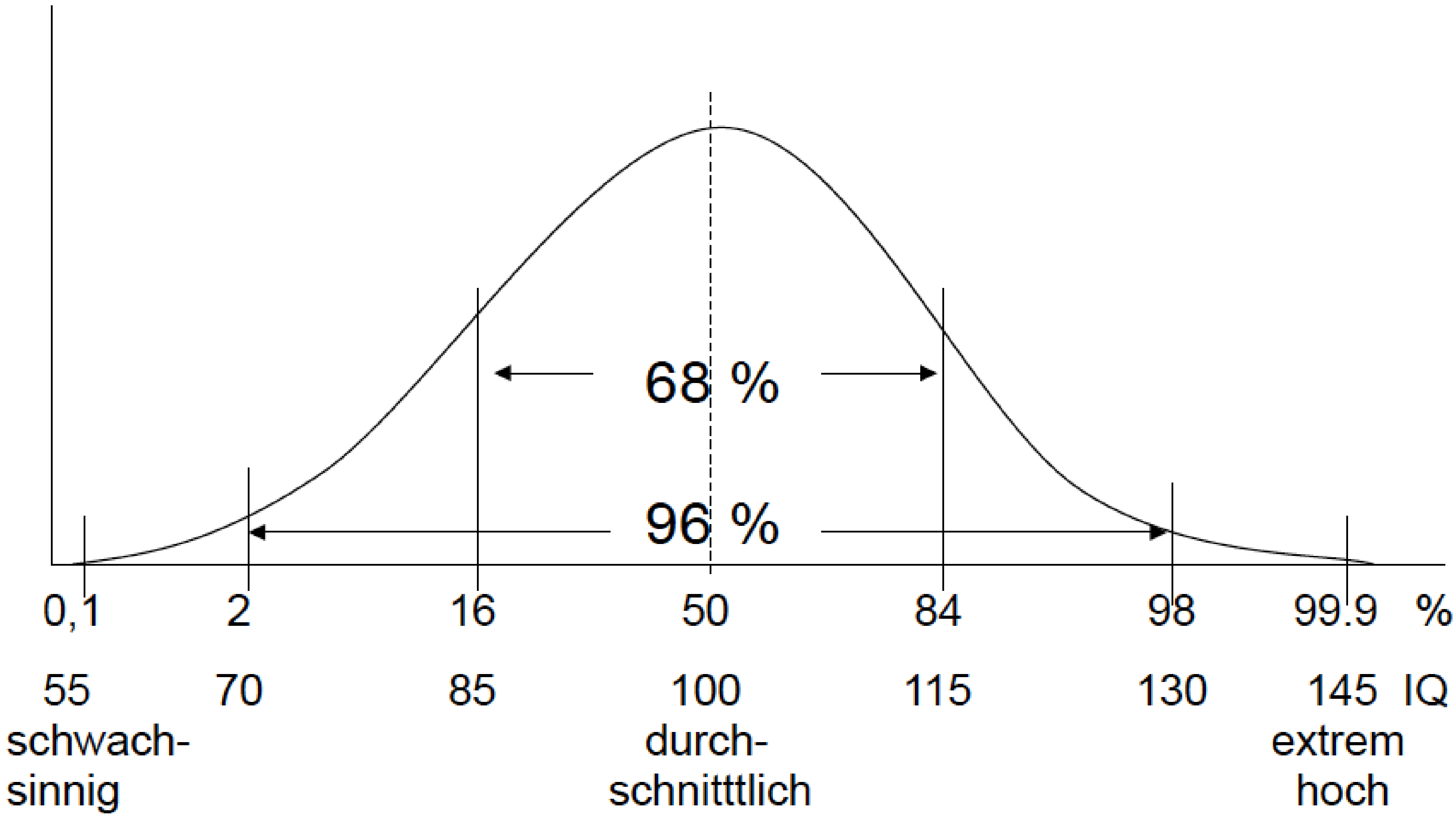
nähert sich asymptotisch 0, wird aber für keinen Wert von x jemals 0

σ bestimmt die Form (steil-flach), μ verschiebt die Kurve entlang der X-Achse

2 Wendepunkte: bei $\mu - \sigma$ und $\mu + \sigma$

Bei ausreichend großem N (etwa > 20) kann die Normalverteilung zur **Approximation der Binomialverteilung** verwendet werden.

Beispiel Intelligenz



(Amelang / Bartussek)

Standardnormalverteilung

Definition

Unter den unendlich vielen Normalverteilungen gibt es eine, die dadurch gekennzeichnet ist, dass sie einen Erwartungswert von $\mu = 0$ und eine Standardabweichung von $\sigma = 1$ hat.

Dieser einfachste Fall der Normalverteilung wird **Standardnormalverteilung** genannt.

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

Eigenschaften

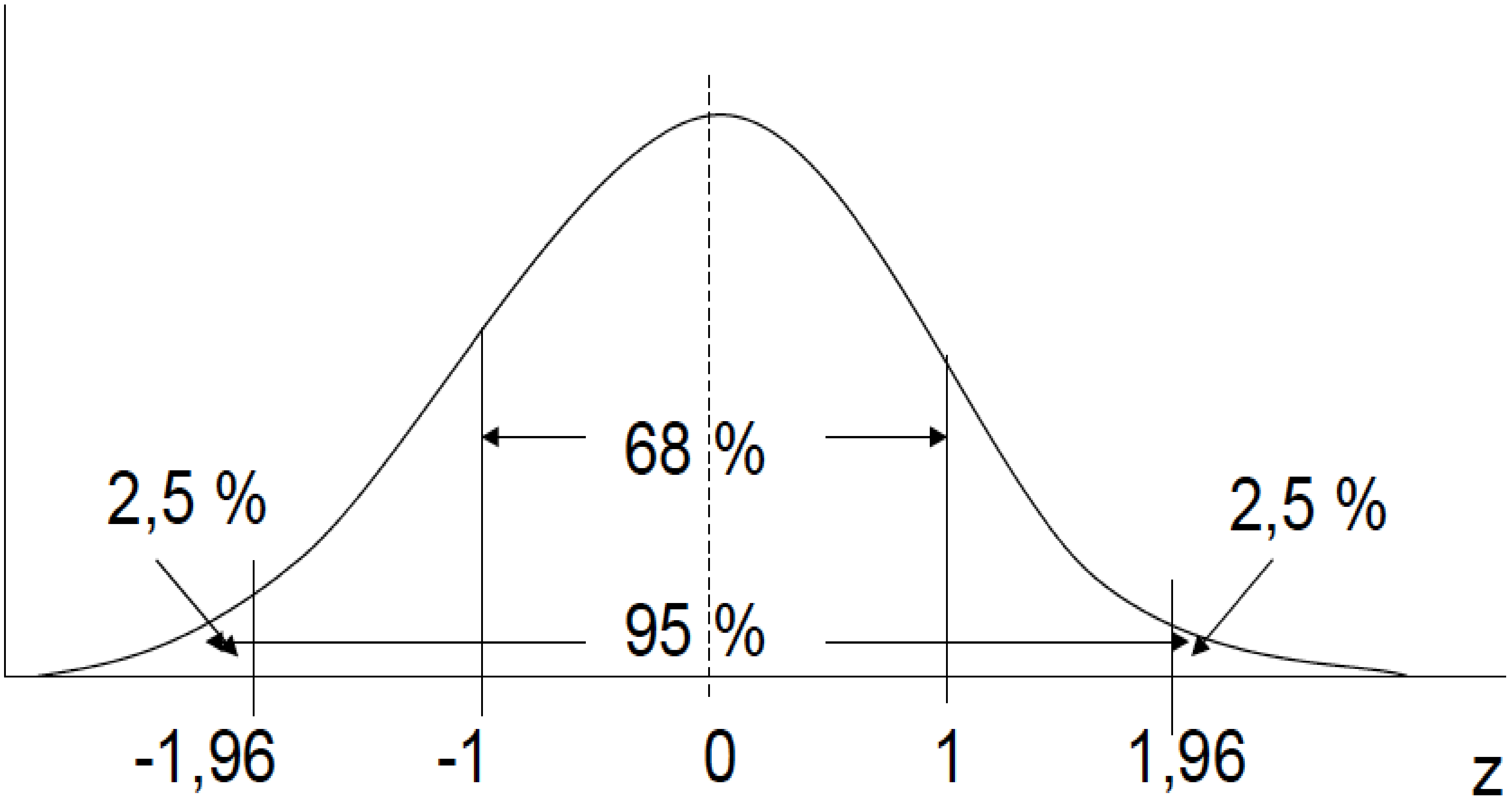
Symmetrieachse bei $x = 0$, Maximum bei $\frac{1}{\sqrt{2\pi}}$

Wendepunkte $x = 1$ und $x = -1$

Jede Normalverteilung lässt sich in eine Standardnormalverteilung überführen.

Z-Transformation

Noch einmal Intelligenz...



Z-Transformation

Definition

Transformation der Werte x_i einer Zufallsvariablen X so, dass anschließend in der neuen Zufallsvariable Z (mit den Werten z_i) $\boldsymbol{\mu = 0}$ und $\boldsymbol{\sigma} = 1$ gilt.

Die Zufallsvariable **Z** ist dann standardnormalverteilt.

Transformation

$$z_i = \frac{x_i - \bar{x}}{s}$$

Von jedem Wert x_i der Zufallsvariable wird der Mittelwert subtrahiert und durch die Standardabweichung dividiert.

Beispiel

$$x_1 = 10,\ x_2 = 11,\ x_3 = 12$$

$$\bar{x} = 11$$

$$s^2 = \frac{1}{n} * \sum_{i=1}^{n} (x_i - \bar{x})^2 = 0,67 \qquad s = \sqrt{s^2} = 0,82$$

$$z_1 = -1,22,\ z_2 = 0,\ z_3 = 1,22$$

$$\bar{x} = 0,\ s^2 = 1,\ s = 1$$

Und noch einmal Intelligenz...

Menschliche Intelligenz (gemessen als IQ) ist in Westeuropa normalverteilt mit einem Durchschnittswert von 100 und einer Standardabweichung von 15.
Wie groß ist die Wahrscheinlichkeit, (hier) einen Menschen einen mit einem IQ von mehr als 130 zu treffen?

Gegeben: $\bar{x} = 100$, $s = 15$

Gesucht: p *(probability)* (x) > 130

(1) Z-Transformation von x = 130

z(130) = (130-100)/15 = 30/15 = 2

$$z_i = \frac{x_i - \bar{x}}{s}$$

(2) Bestimmung der Wahrscheinlichkeit p

Tabelle der Standardnormalverteilung

$p(z \leq 2)$ - >Tabelle = .9772 = 97,72 %

$p(z > 2)$ = 2,28 Prozent

(3) Ablesen – und fertig.

Wir erinnern uns...

Bei stetigen Zufallsvariablen gilt:

$$p\ (a \leq x \leq b) = \int_a^b f(x)\,dx$$

Die Flächeninhalte unter dem Graphen für die Standardnormalverteilung berechnen sich aus:

$$\Phi_{0;1}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}x^2}\,dx$$

Tabelle der Standardnormalverteilung

Aus der Tabelle kann die Wahrscheinlichkeit $\Phi(z)$ für die Standardnormalverteilung abgelesen werden.

Aus Gründen der Symmetrie enthält die Tabelle keine negativen Werte.

Es gilt: $\Phi(-z) = 1 - \Phi(z)$

Z-Tabelle Teil 1

$\phi_{0;1}(z)$

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774

Z-Tabelle Teil 2

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736

Z-Tabelle Teil 3

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

Zufallsstichprobe

Definition

Wenn jedes Element der Grundgesamt die gleiche Chance hat, in die Stichprobe aufgenommen zu werden, handelt es sich um eine **Zufallsstichprobe**.

Mit Hilfe der Wahrscheinlichkeitstheorie ist es dann möglich, von den Stichprobenergebnissen auf die entsprechenden Parameter der Grundgesamtheit zu schließen, d.h. **diese zu schätzen**.

Inferenzschluss

Intervallschätzung und Konfidenzintervall

Definition

Schätzt man einen unbekannten Parameter in der Grundgesamtheit (z.B. μ) mit Hilfe eines Stichprobenkennwertes (z.B. $\bar{x}$), so **trifft** man den Parameter **niemals genau** (Problem der Punktschätzung).

Mit Hilfe des **Konfidenzintervalls (KI)** lässt sich jedoch eine Aussage darüber machen, in welchem Bereich sich der unbekannte Parameter mit einer **vorgegebenen Wahrscheinlichkeit** befindet.

Vertrauenswahrscheinlichkeit

Die Stichprobenkennwerte sind **Zufallszahlen.** Deren Verteilung ist bekannt: **Normalverteilung** (Zentraler Grenzwertsatz).

Schätzfunktion

Allgemeine Form

$\bar{x} \rightarrow \mu$ $\quad KI = \bar{x} \pm z * \frac{s}{\sqrt{n}}$ **Mittelwertschätzung**

$p \rightarrow \pi$ $\quad KI = p \pm z * \sqrt{\frac{p * (1-p)}{n}}$ **Prozentanteilsschätzung**

Eigenschaften

Das Konfidenzintervall **vergrößert** sich...

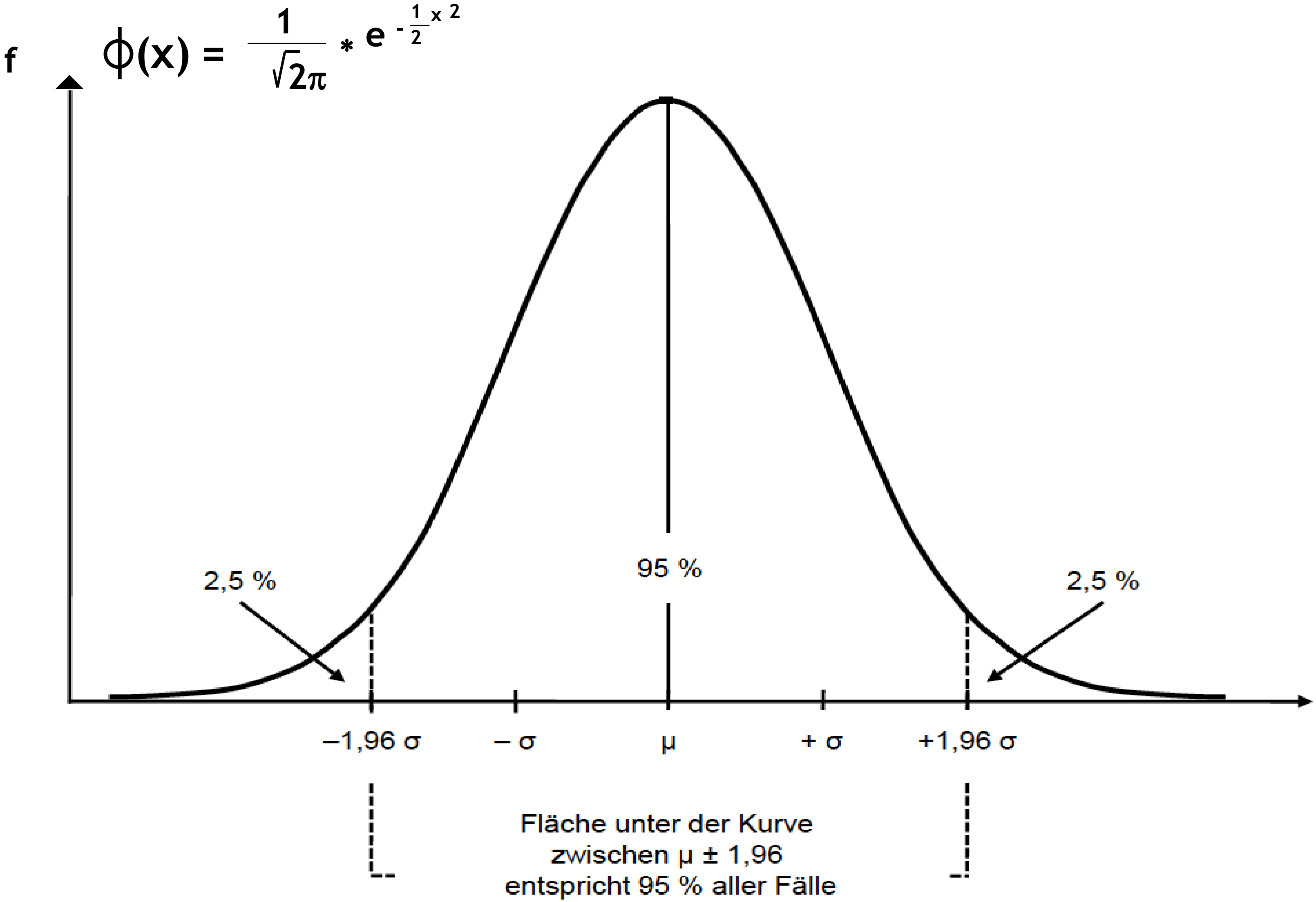
...wenn der **Standardfehler** **größer** wird, d.h. wenn die **Fallzahl** **sinkt**, bzw. wenn die **Streuung** in der Population **größer** wird.

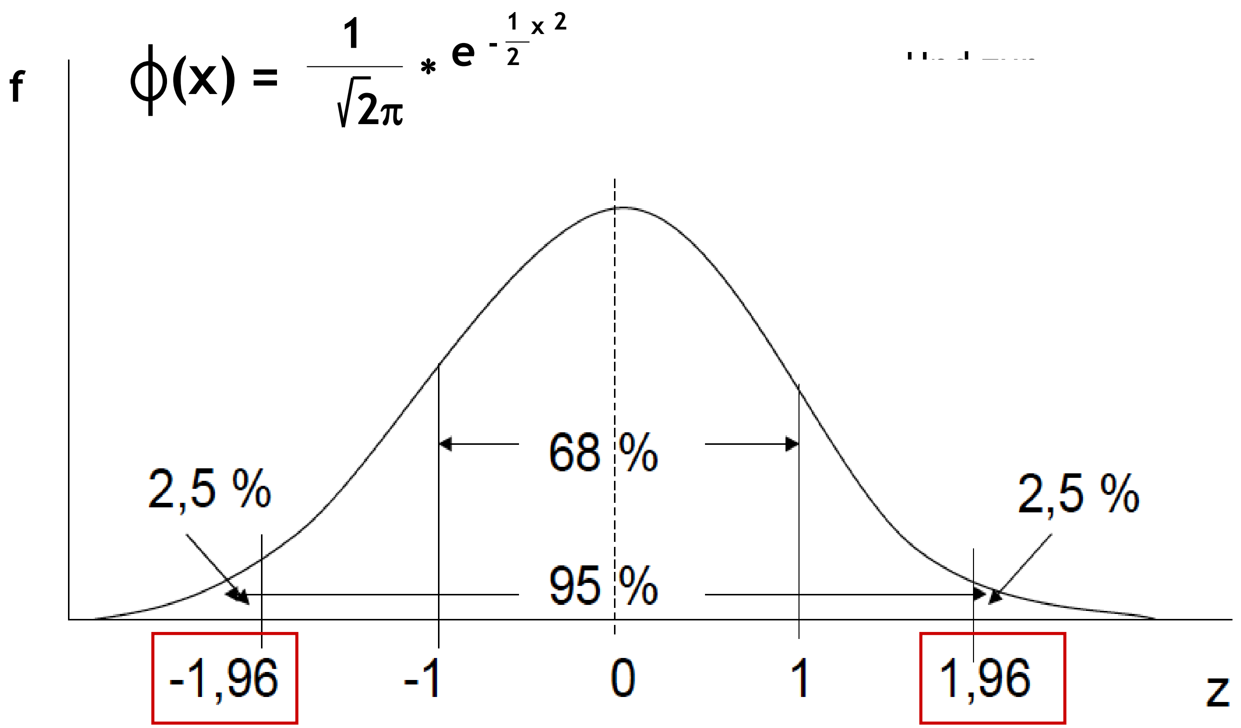
... wenn die **Vertrauenswahrscheinlichkeit** **erhöht** wird (z.B. von 95% auf 99%)

$$KI = \bar{x} \pm \boxed{z} * \boxed{\frac{\boxed{s}}{\sqrt{\boxed{n}}}}$$

(Beispiel: KI für Mittelwert)

Graphik Standardnormalverteilung





Wie kommt man bloß auf diese 1,96?

(1) Festlegung der gewünschten Vertrauenswahrscheinlichkeit (z.B. 95%, p=.95). Das gesuchte Intervall soll also in 95 von 100 Fällen den Parameter der Grundgesamtheit (hier: μ) enthalten.

Gesucht:

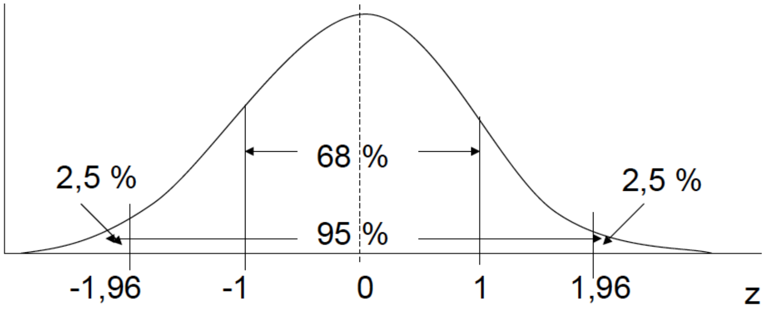
$$\mathbf{KI}\ (\mu_{95}) = \bar{x} - a \leq \mu \leq \bar{x} + a$$

(2) Da die Fläche unter der Kurve immer gleich 1 ist, gilt für die Rest- bzw. Gegenwahrscheinlichkeit 1 - 0.95 = 0.05, aufgrund der Symmetrie der Kurve also genauer gesagt: beidseitig 0.025.

Anders formuliert:

100% - 95% = 5% = 2 x 2,5%

Für die Standardnormalverteilung und eine Vertrauenswahrscheinlichkeit von p = 95% benötigt man dieses 2,5% Quantil:



$$z_{\frac{1+p}{2}} = z_{\frac{1+.95}{2}} = z_{.975} = \pm 1.96$$

z-Tabelle again

$\phi_{0;1}$ **(z)**

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736

Herleitung

(3) **Zentraler Grenzwertsatz**

z liegt mit einer Wahrscheinlichkeit von **p** zwischen den angegebenen Intervallgrenzen der Standardnormalverteilung.

$$P\left(-z_{\frac{1+p}{2}} \leq z_{\bar{x}} \leq +z_{\frac{1+p}{2}}\right) = p$$

Zum Beispiel:

$P(-19.6 \leq z \leq +1.96) = 0.95 = 95\%$

(4) z ist das Ergebnis der Standardisierung:

$$z = \frac{\bar{x} - \mu}{\boxed{\frac{s_x}{\sqrt{n}}}}$$

Standardfehler des Mittelwerts (SE)

(5) Einsetzen und nach **μ auflösen:**

μ liegt mit der angegebenen Wahrscheinlichkeit **(γ) ± z ∗ SE** vom Mittelwert entfernt.

$$P\left(-z_{\frac{1+p}{2}} \leq \frac{\bar{x} - \mu}{\frac{s_x}{\sqrt{n}}} \leq +z_{\frac{1+p}{2}}\right) = p$$

$$P\left(\bar{x} - z_{\frac{1+p}{2}} \frac{s_x}{n} \leq \mu \leq +\bar{x} + z_{\frac{1+p}{2}} \frac{s_x}{n}\right)$$

Fragestellungen

Wie ist das arithmetische Mittel in der Grundgesamtheit?
Unterscheiden sich die Mittelwerte zweier oder mehrerer Gruppen?
In welchem Bereich liegt ein bestimmter Prozentsatz meiner Werte?

Voraussetzungen

Messwerte sind metrisch (oder quasi-metrisch) skaliert.
Messwerte sind normalverteilt.
Stichprobengröße sollte $n \geq 30$ sein.

Berechnung

Konfidenzintervall für Mittelwerte

$$\bar{x} \rightarrow \mu \qquad KI = \bar{x} \pm z * \boxed{\frac{s}{\sqrt{n}}}$$

Standardfehler (geschätzt aus Werten der Stichprobe)

Konfidenzintervall für Prozentwerte

$$p \rightarrow \pi \qquad KI = p \pm z * \sqrt{\frac{p * (1-p)}{n}}$$

Gängige z-Werte

95 % = $\pm$ 1.96   99 % = $\pm$ 2.58   99,9 % = $\pm$ 3.29

Kleiner Exkurs:

Wann und warum darf man den Standardfehler aus der Stichprobe schätzen?

Grundgesamtheit, Stichprobe, Varianz

Die unbekannte Varianz in der Grundgesamtheit σ^2 wird mit Hilfe der Stichprobenvarianz s^2 geschätzt.

Grundgesamtheit

tatsächliche Werte

$$\boxed{\sigma^2} = \frac{1}{\boxed{n}} * \sum_{i=1}^{\boxed{n}} (x_i - \bar{x})^2$$

Stichprobe

geschätzte Werte

$$\boxed{s^2} = \frac{1}{\boxed{n-1}} * \sum_{i=1}^{\boxed{n}} (x_i - \bar{x})^2$$

Da aber die Stichprobenvarianzen kleiner sind als die Varianz der Grundgesamtheit, bedarf es eines Korrekturfaktors.

Für die Standardabweichung gilt:

$$\sigma = s = \frac{s}{\sqrt{n}}$$

Standardfehler

Standardfehler

Definition

Der Standardfehler ist ein Maß für die Genauigkeit einer Messung: je geringer, desto genauer.

Anders gesagt:

Er gibt an, welcher Fehler bei der Schätzung zu erwarten ist.

Eigenschaften

Der Standardfehler hängt proportional von der realen Streuung in der Grundgesamtheit ab.

Er ist umgekehrt proportional zu $\sqrt{n}$.

Beispiel

n = 400; Mediennutzung in Minuten,
Mittelwert 17,8, Standardabweichung 3,6

$$\frac{3{,}6}{\sqrt{400}} = \frac{3{,}6}{20} = \mathbf{0{,}18}$$

$$s = \frac{s}{\sqrt{n}}$$

n = 100; Mediennutzung in Minuten,
Mittelwert 17,8, Standardabweichung 3,6

$$\frac{3{,}6}{\sqrt{100}} = \frac{3{,}6}{10} = \mathbf{0{,}36}$$

Anwendungsbeispiel 1

In einem Methodenkurs mit n=100 Teilnehmern wurde eine durchschnittlicher IQ von 102 gemessen. Die Standardabweichung beträgt 15.

Wie groß ist das (95%)-Konfidenzintervall des gemessenen Mittelwertes? Sind die Studierenden des Methodenkurses mit über-(oder unter-) durchschnittlicher Intelligenz ausgestattet?

Lösung

Gegeben:

$n = 100$

$\bar{x} = 102$

$s = 15$

Gesucht: KI (95%)

$$\bar{x} \rightarrow \mu \qquad KI = \bar{x} \pm z * \frac{s}{\sqrt{n}}$$

$$\rightarrow z(95\%) = 1.96$$

$$KI\ (95\%) = 102 \pm 1{,}96 * 15/10$$

$$= 102 \pm 2{,}94$$

$$\boxed{99{,}06 \leq \mu \leq 104{,}95}$$

In einer n = 1.000 Stichprobe beträgt der Anteilswert der beschäftigten Frauen 42% (p = .42).

Wie lautet das entsprechende (95%) Konfidenzintervall für diesen Anteilswert?

Gesucht: KI(95%)

$$p \rightarrow \pi \qquad KI = p \pm z * \sqrt{\frac{p * (1-p)}{n}}$$

Lösung

z (95%) = 1.96

Der Stichprobenfehler bei n = 1.000 beträgt:

$$\sqrt{\frac{p * (1 - p)}{n}} = \sqrt{\frac{0{,}42 * 0{,}58}{1.000}} = 0{,}0156$$

$$KI(95\%_l) = 0{,}42 - 1{,}96 * 0{,}0156 = 0{,}3894$$

$$KI(95\%_r) = 0{,}42 + 1{,}96 * 0{,}0156 = 0{,}4506$$

$$\boxed{38{,}9\% \leq p \leq 45{,}1\%}$$

Beispiel Intervallschätzung

n = 100 repräsentative Bevölkerungsstichproben

Mediennutzung pro Tag in Minuten

Rote Linie:
tatsächliches (aber unbekanntes) AM in der Grundgesamtheit (GG)

Waagerechte Linien:
Intervallschätzung (Konfidenzintervalle) für AM in der jeweiligen Stichprobe

gewählte Vertrauenswahrscheinlichkeit 95%

= in 5 von 100 Fällen liegt das AM der GG trotzdem außerhalb des Intervalls

(blaue Linien)

Ausführlicher ins Thema einsteigen:

Zum Beispiel:

https://de.wikiversity.org/wiki/Kurs:Stochastik/Normalverteilung