

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

QSB1 - Tag 3

5. Zufallsvariablen / Verteilungen

Dr. rer. nat. Meinhard Mende,

IMISE, Universität Leipzig

imise.

Lernziele

Nach Beendigung dieser Lehreinheit

- kennen Sie die Definition von Zufallsvariablen,
- können Sie die Binomialverteilung für Berechnungen anwenden,
- ist Ihnen die Normalverteilung mit ihren Eigenschaften vertraut,
- können Sie die z-Transformation durchführen und
- Wahrscheinlichkeiten aus Verteilungstabellen ablesen.

V 5

5.0 Grundlagen und Konzepte

Einführung

Bisher: Beschreibung empirisch gewonnener Daten in einer Stichprobe

Frage:

- Was lässt sich aus diesen Beobachtungen schlussfolgern?
  (z. B.: Kann ich eine Schätzung für den wahren Wert gewinnen?
  Wie genau ist sie? Ist sie mit einer bestimmten Erwartung vereinbar?)

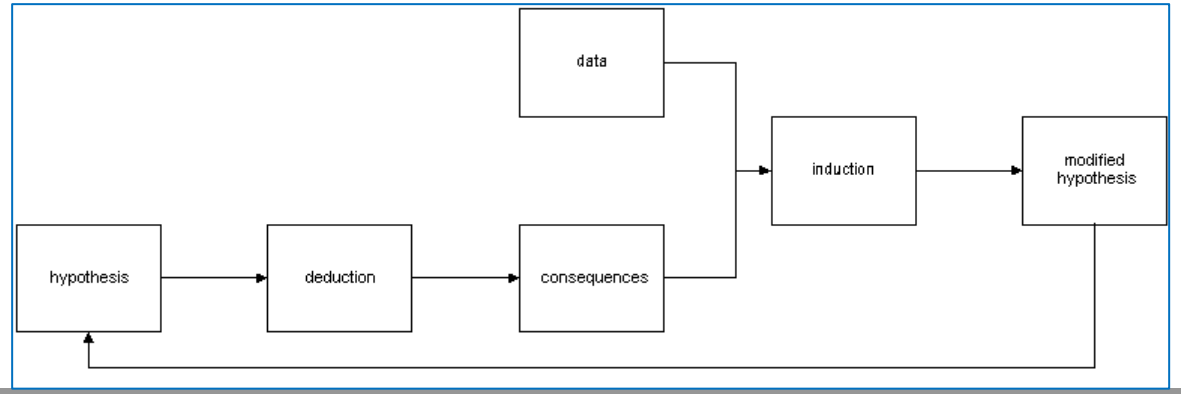
- Neuer Ansatz: für statistische Schlussfolgerungen ein theoretisches Modell zur Verteilung von Merkmalsausprägungen anzunehmen und die Daten von Stichproben auf dieser Basis auszuwerten.

Wir denken in Modellen

- *We all depend on models to interpret our everyday experiences. We interpret what we see in terms of mental models constructed on past experience and education. They are constructs that we use to understand the pattern of our experiences.*
  David Bartholomew (J Royal Stat Soc A 1995; 158(1): 1-20)

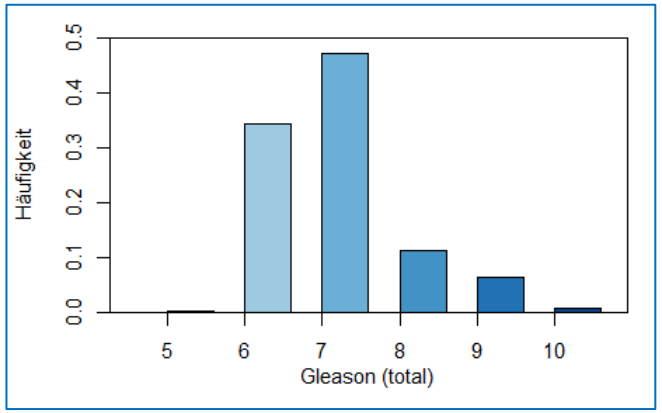
- *All models are wrong, but some are useful.*
  George E. P. Box 1976

- Chance: Components of Statistical Thinking …
  Journal of Statistics Education 2002;10:3



Rückblick: Relative Häufigkeiten

Gleason sum (preop), *n* (%)	
5	1 (0.2)
6	188 (34)
7	258 (47)
8	61 (11)
9	35 (6.4)
10	4 (0.7)



- In V3 haben wir über den Werten im Wertebereich 5 – 10 die relativen Häufigkeiten aufgetragen.
- Relative Häufigkeiten (aus Daten) entsprechen in der Theorie Wahrscheinlichkeiten.
- Diese Wahrscheinlichkeiten $p_1, p_2, \dots, p_k$ bilden über dem Wertebereich $(x_1, x_2, \dots, x_k)$ die Wahrscheinlichkeitsverteilung.

Zufallsvariablen

- Zufallsvariable: numerisches Ergebnis eines Zufallsexperimentes
- Die Verteilung ihrer Werte kann durch eine (theoretische) Verteilungs-funktion beschrieben werden.

Beispiel 5.1: Zufallsexperiment: Münzwurf mit zwei Münzen

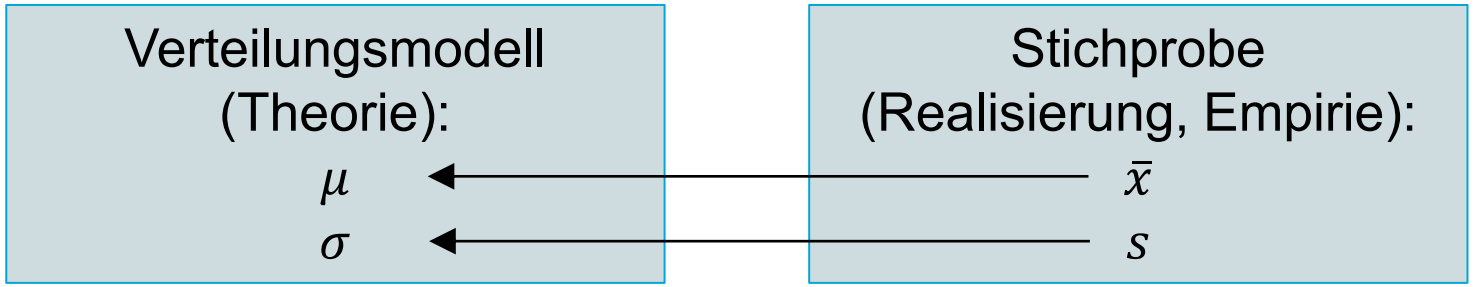
Zufallsvariable *X:* Anzahl von Wappen

Outcomes:	ZZ	ZW	WZ	WW
				↘ ↙
Realisierungen von $X \rightarrow x$:	0		1	2

- Die Wahrscheinlichkeit, dass die Zufallsvariable X den Wert x annimmt, schreibt man als $P(X = x)$ oder $p(x)$.
- z. B.: $P(X = x)$: ¼ ½ ¼

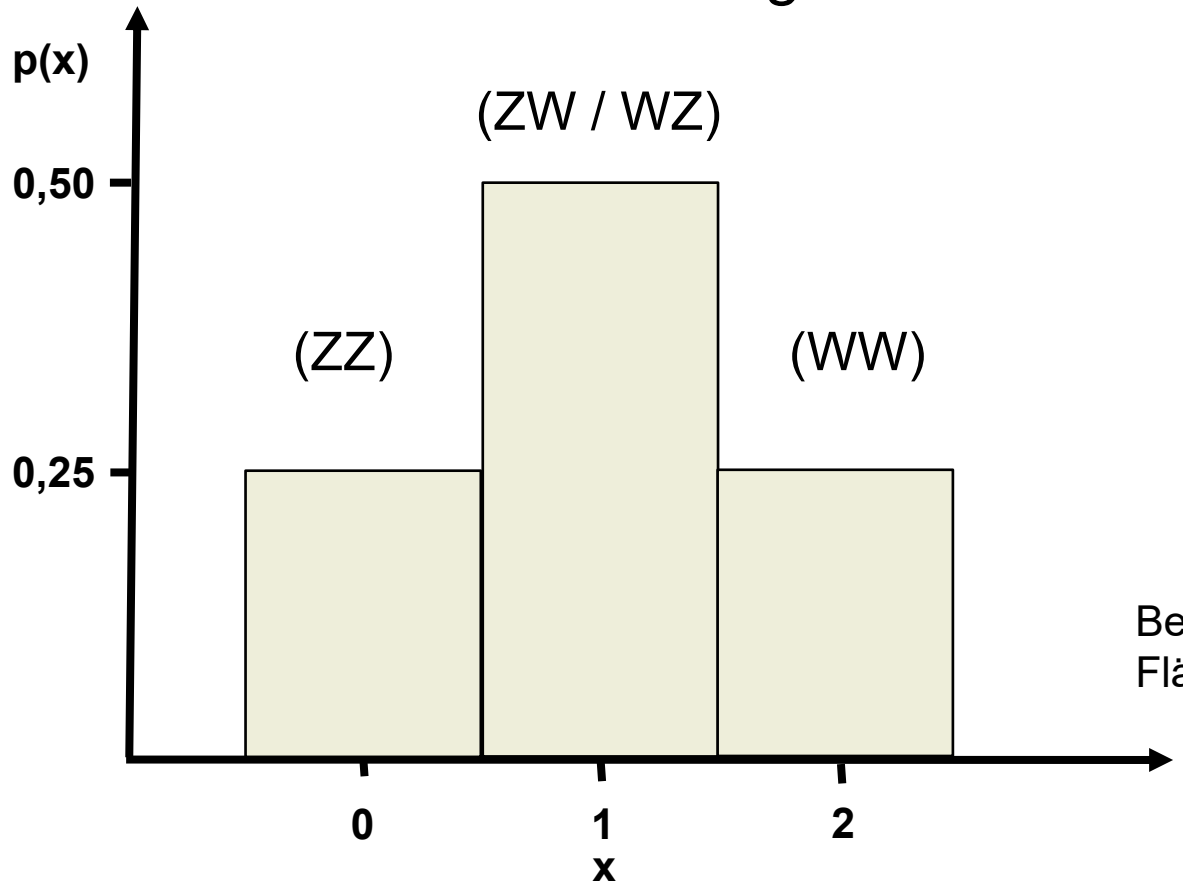
Zufallsvariablen II

- Je nach Wertebereich unterscheidet man diskrete und stetige Zufallsvariablen:
  - diskrete Zufallsvariablen: Wertebereich endlich viele, diskrete $x_1, x_2, \ldots, x_k$
  - stetige Zufallsvariablen: Wertebereich ist ein ganzes Intervall (Kontinuum) der Zahlengeraden.
- Notation:
  - lateinische Buchstaben: Daten aus der Stichprobe
  - griechische Buchstaben: Parameter (Kenngrößen) einer Wahrscheinlichkeitsverteilung



B 5.1 Wahrscheinlichkeitsverteilung

Histogramm der Wahrscheinlichkeitsverteilung:



Bemerke:
Flächeninhalt = 1

Erwartungswert

- Der Erwartungswert einer diskreten Zufallsvariable ist definiert als:

$$E(X) = \mu = \sum_j x_j \cdot p_j,$$

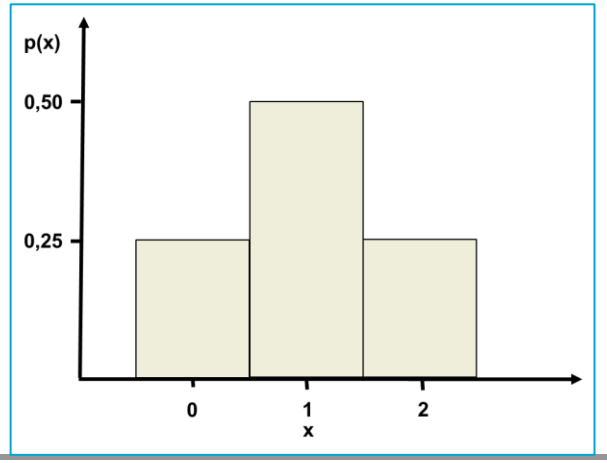
- d. h. jeder Wert wird mit der Wahrscheinlichkeit gewichtet, mit welcher er auftritt.



Gonick & Smith: Statistik in Cartoons

- zum Beispiel B 5.1:

$$\mu = 0 \cdot 0{,}25 + 1 \cdot 0{,}50 + 2 \cdot 0{,}25 = 1$$



Erwartungswert und Varianz I

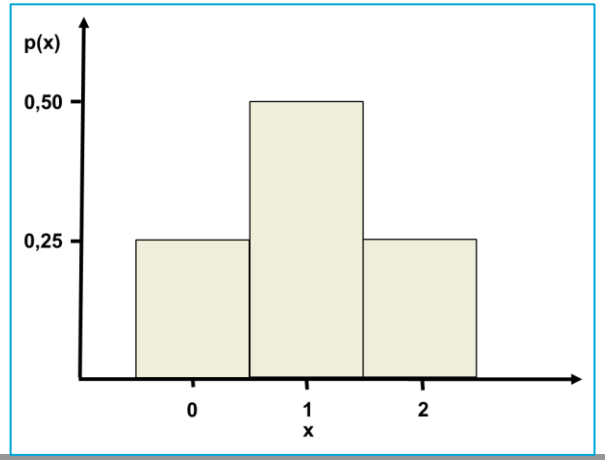
Die Varianz einer diskreten Zufallsvariable ist definiert als:

$$V(X) = \sigma^2 = E[(X-\mu)^2] = \sum_j \left(x_j - \mu\right)^2 \cdot p_j$$

- D. h., das ist die erwartete quadrierte Abweichung der Zufallsvariablen *X* von ihrem Erwartungswert *E(X)*.

- zum Beispiel B 5.1:

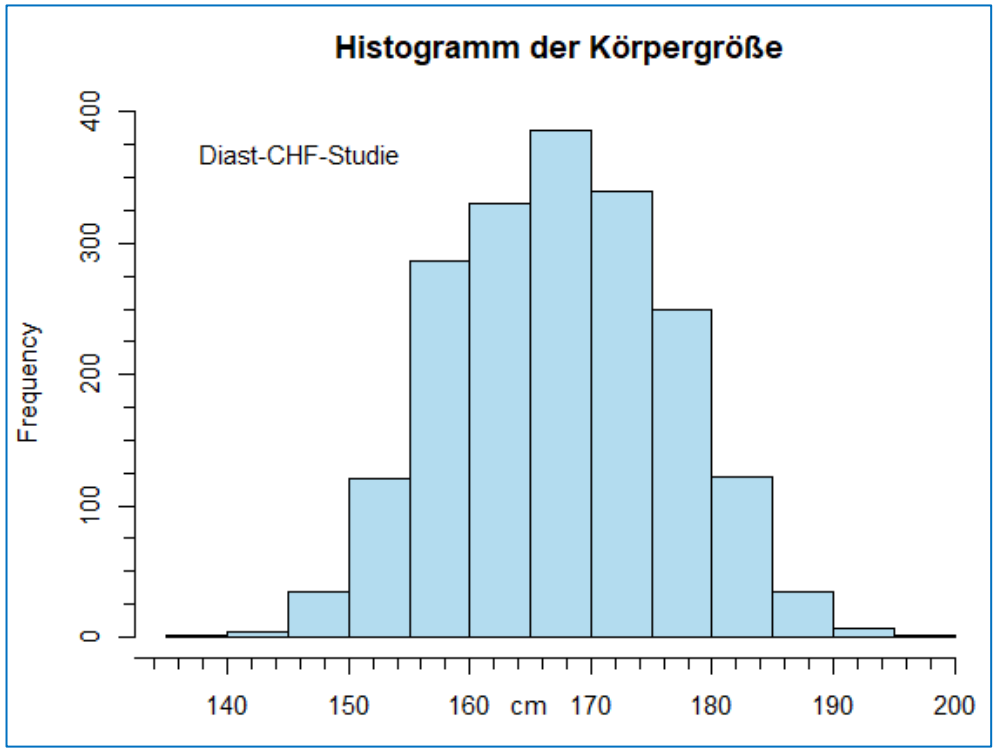
$$\sigma^2 = (0-1)^2 \cdot 0{,}25 + (1-1)^2 \cdot 0{,}50 + (2-1)^2 \cdot 0{,}25 = 0{,}50$$



Stetige ZV: Histogramm

B 5.2: Diast-CHF-Studie,
Beobachtungsstudie: $n_1 = 1727$ Probanden mit diastolischer Herzinsuffizienz bzw. Risikofaktoren dafür und $n_2 = 202$ Kontrollen,

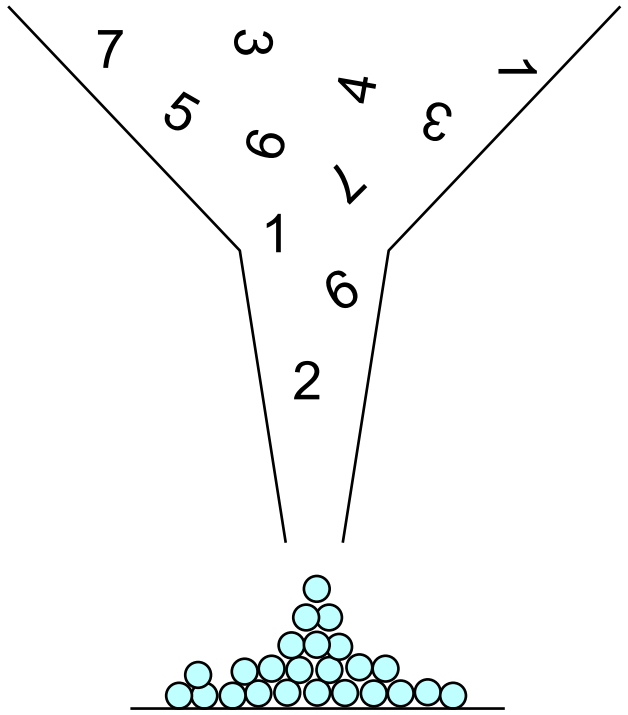
Follow-up: 10 Jahre
Variable Körpergröße baseline



- Alter ist eine kontinuierliche (stetige) Zufallsvariable, d. h.
- hier sind im Intervall [130, 200] (theoretisch) alle Werte möglich.
- Werte kommen in der Mitte gehäuft vor.

Modell: Verteilungsdichte

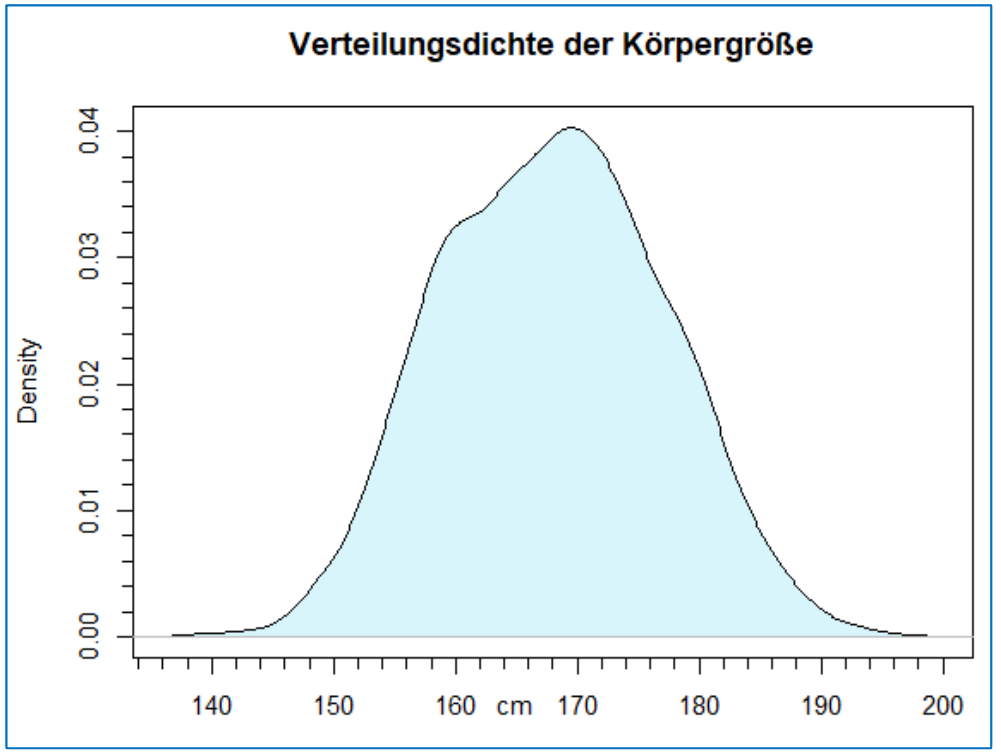
- Werte der Stichprobe häufen sich auf der x-Achse und bilden die Verteilungsdichte
- Diese wird durch Vorgabe von Bedingungen modelliert
- und es wird eine Formel entwickelt (z. B. Normalverteilung).
- Die Verteilungsdichte repräsentiert Wahrscheinlichkeiten, also ergibt ihre Gesamtsumme 1.



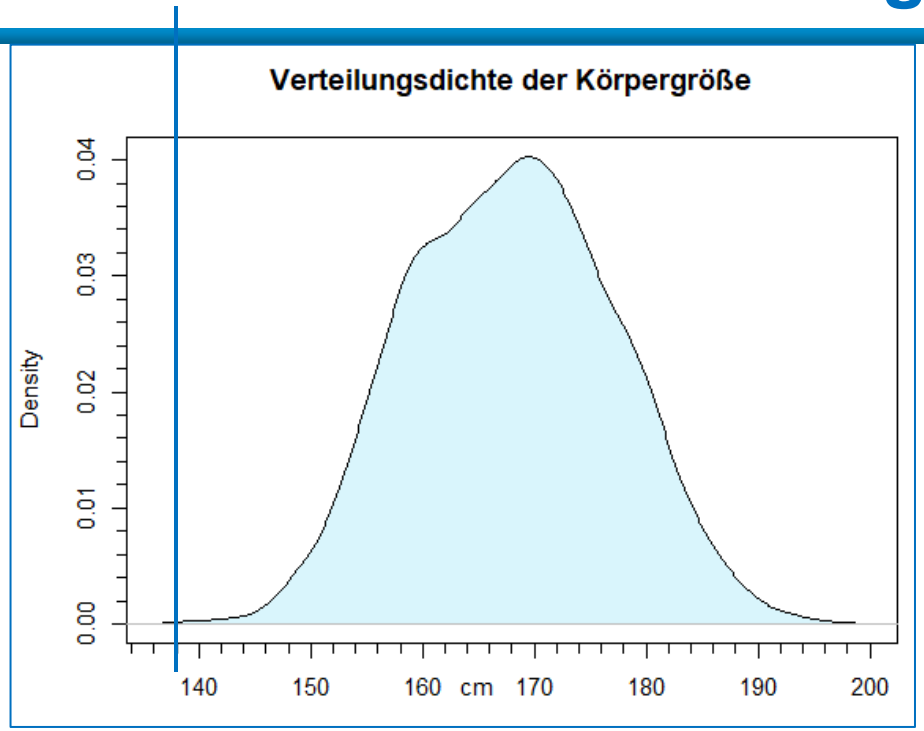
Stetige ZV: Verteilungsdichte

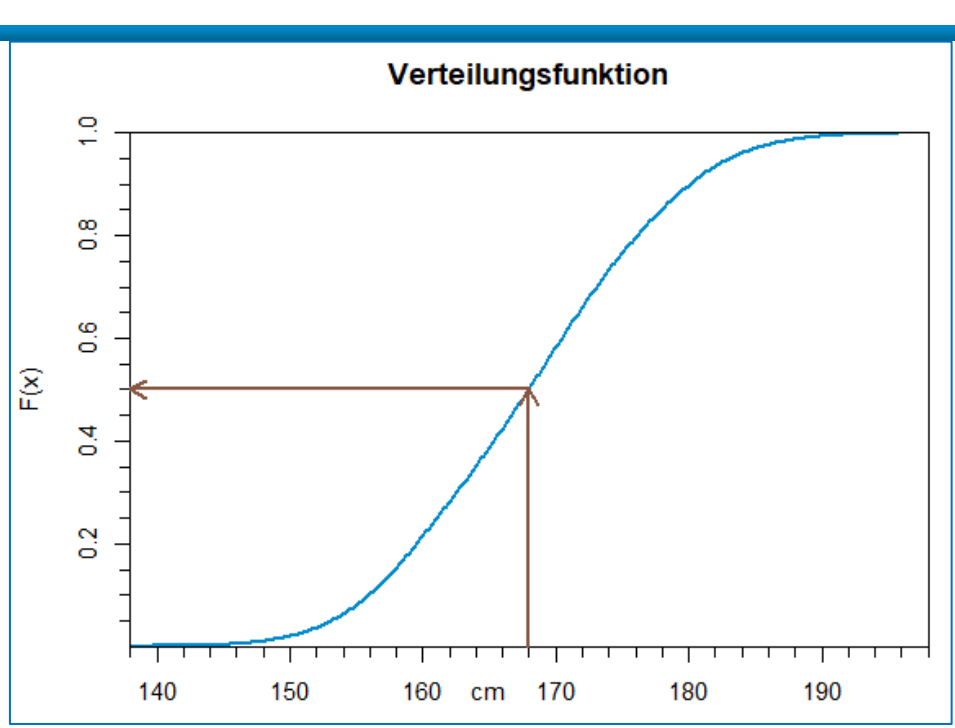
B 5.2: Diast-CHF-Kohorte; Körpergröße baseline

- Die diskreten Säulen des Histogramms verschmelzen im Grenzwert zu einer Kurve.
- Die Fläche unter der Kurve ist (per Def.) 1.



Verteilungsfunktion





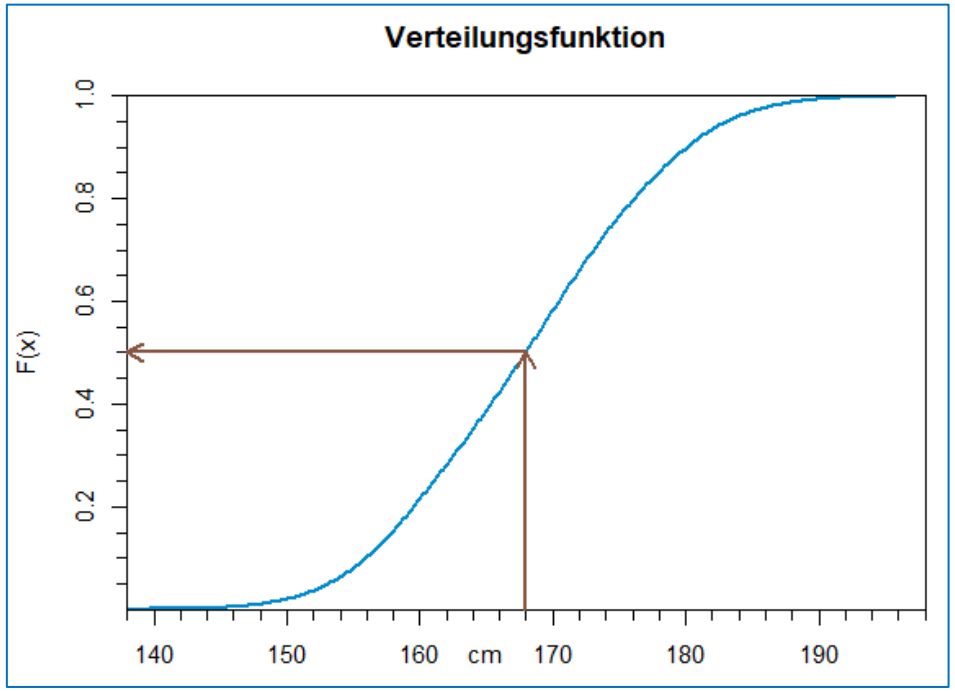
Die Verteilungsfunktion F(x) gibt an, mit welcher Wahrscheinlichkeit ein Wert $\leq x$ erzielt wird: $F_X(x) = P(X \leq x)$.

- Sie ergibt sich durch Integration der Verteilungsdichte von links bis x; d. h.,
- sie stellt die Fläche unter der Kurve links der Linie bei x dar.

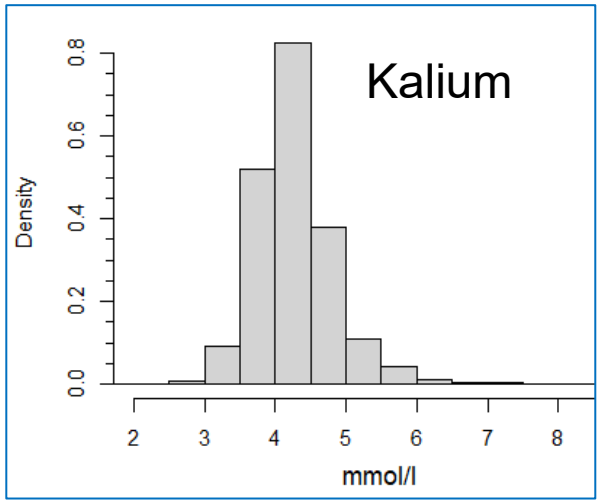
Eigenschaften der Verteilungsfunktion

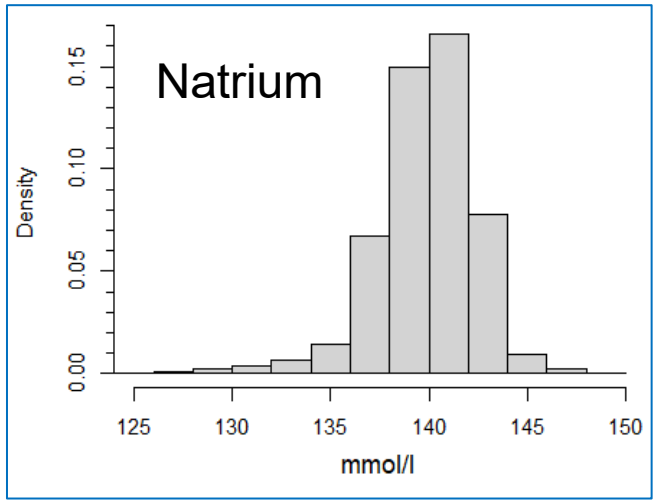
Ihr y-Wert ist

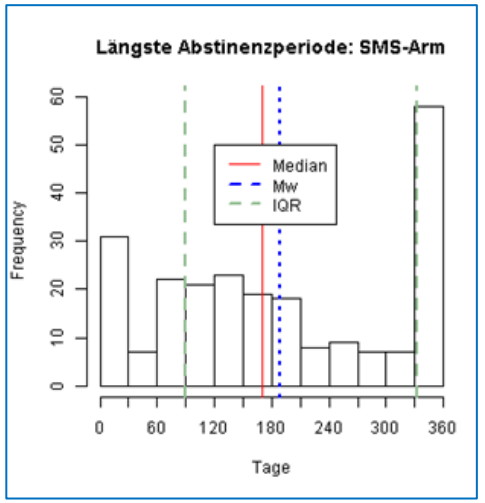
- unterhalb des Minimums: = 0,
- oberhalb des Maximums: = 1;
- Sie wächst monoton.



Verschiedene empirische Verteilungen

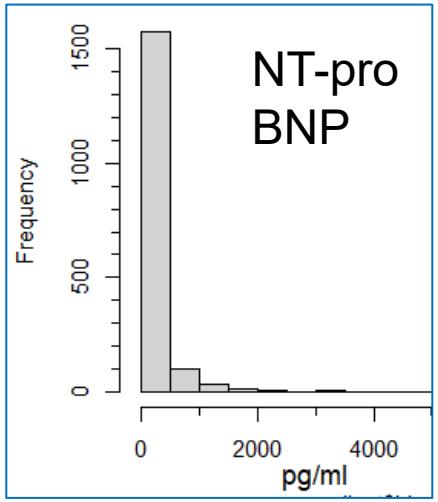






Verteilungen unterscheiden sich hinsichtlich

- Wertebereich
- Symmetrie bzw. Schiefe
- extremen Werten



Erwartungswert und Varianz II

Der Erwartungswert einer stetigen Zufallsvariable ist definiert als:

- $E(X) = \mu = \int x \cdot f(x)\,\mathrm{d}x$
- Dabei bezeichnet $f(x)$ die Wahrscheinlichkeitsdichte.
- Der Erwartungswert markiert den Schwerpunkt einer Verteilung.

Die Varianz einer stetigen Zufallsvariable ist definiert als:

- $V(X) = \sigma^2 = E[(X - \mu)^2] = \int (x - \mu)^2 \cdot f(x)\,dx$
- Die Standardabweichung ist die Wurzel der Varianz.

Lineare Funktionen von ZV

Ist Y eine lineare Funktion einer Zufallsvariablen X. Dann gilt:

- $E[Y] = E[a \cdot X + b] = a \cdot E[X] + b$
- $V[Y] = V[a \cdot X + b] = a^2 \cdot V[X]$

Beispiel 5.2: Sie setzen 55 € und gewinnen 100 €, falls beim Münzwurf Zahl fällt. Wie hoch ist der erwartete Gewinn?

X – Zufallsvariable Münzwurf [0 = Wappen, 1 = Zahl], dann ist:

- $p(x) = 0{,}5;$
- $E[X] = 0 \cdot 0{,}5 + 1 \cdot 0{,}5 = 0{,}5$
- Erwarteter Gewinn:

$Gewinn =$

$E[G] =$

Summe von ZV

Ist Y die Summe der Zufallsvariablen X_1 und X_2:

- $Y = X_1 + X_2$,
- dann ist Y wieder eine Zufallsvariable.

Und es gilt für Erwartungswert und Streuung von Y:

- $E[X_1 + X_2] = E[X_1] + E[X_2]$
- $V[X_1 + X_2] = V[X_1] + V[X_2] + 2 \cdot Cov[X_1, X_2]$



Sind X_1 und X_2 unabhängig, gilt

- $V[X_1 + X_2] = V[X_1] + V[X_2]$

Spezielle Verteilungen

Wichtige Verteilungsfunktionen in der Wahrscheinlichkeitsrechnung sind:

1. die Binomialverteilung binärer Merkmale
2. die Normalverteilung stetiger Merkmale

Die Normalverteilung wird auch als Gaußverteilung bezeichnet.

V 5

5.1 Binomialverteilung

Modellierung binärer Ereignisse

Wir betrachten einen Zufallsprozess mit genau zwei möglichen Ausgängen.

- Den uns interessierenden Ausgang (Outcome) bezeichnen wir als Erfolg (meist codiert durch 1),
- den uns weniger interessierenden als Misserfolg (codiert durch 0)

Beispiele: + Kopf (= Erfolg) oder Wappen beim Münzwurf,

+ tumorfreier Schnittrand R0 (= Erfolg) oder Tumorinfiltration R1

Bernoulli-Kette

Modellierung:

- Ein solches Zufallsexperiment heißt Bernoulli-Versuch (Jakob I Bernoulli 1654 - 1705) falls es folgende Eigenschaften erfüllt:

1. Das Experiment hat genau 2 Ausgänge: Erfolg (1) oder Misserfolg (0).
2. Die Wahrscheinlichkeit p für einen Erfolg ist in jedem Versuch dieselbe.
3. Die Versuche sind unabhängig: Der Ausgang eines Versuchs hat keinen Einfluss auf spätere Ausgänge.

Hinweis:

- Beim Tumor-Beispiel (letzte Folie) abstrahiert man von den vielen Begleitumständen, welche jede Wiederholung etwas anders gestalten.

Binomialverteilung I

- Durch n-fache Wiederholung des Bernoulli-Versuches mit der Erfolgswahrscheinlichkeit p erzeugen wir eine neue Zufallsvariable.
- Diese binomialverteilte Zufallsvariable X beschreibt die Anzahl an Erfolgen in n Bernoulli-Versuchen mit Erfolgswahrscheinlichkeit p;
- d. h. X ist die Summe der Ergebnisse von n realisierten Bernoulli-Versuchen

z. B.: X : Anzahl von Wappen (Erfolg) in einem zweifachen Münzwurf

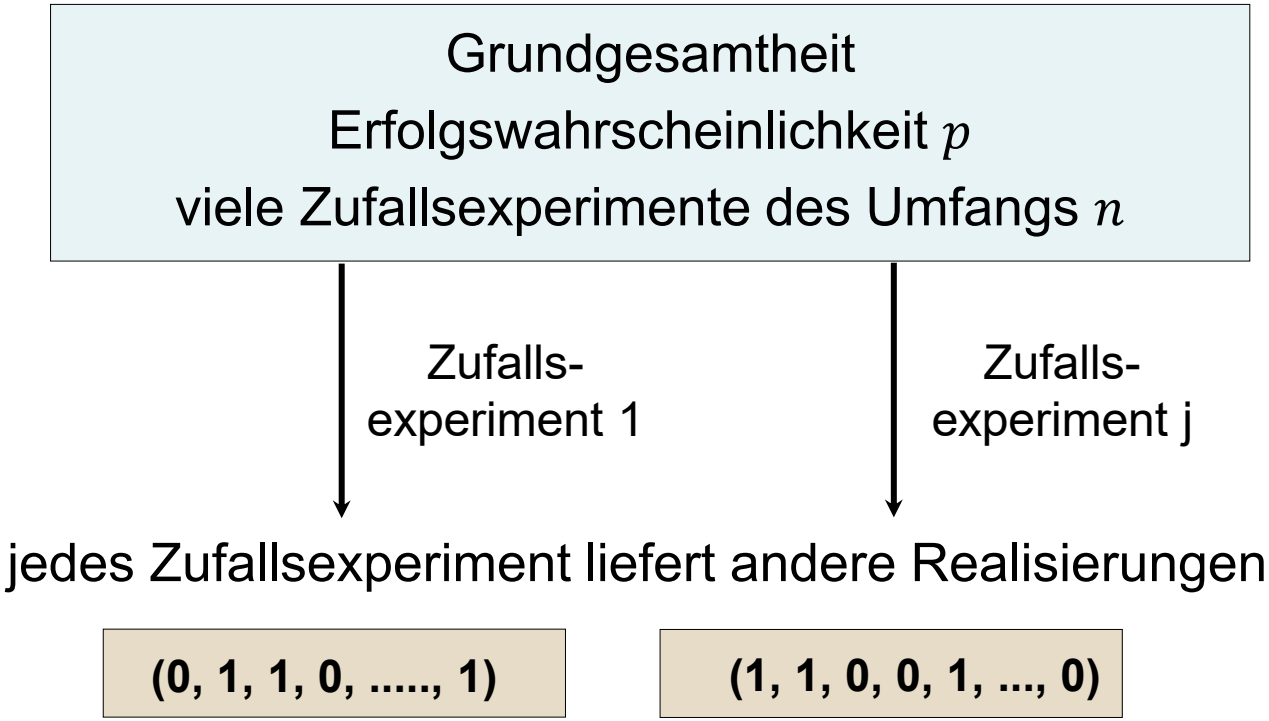
$$n = 2, p = 0{,}5$$

$$k = \text{Anzahl an Erfolgen}$$

k:	0,	1,	2
$P(X = k)$:	¼	½	¼

Binomialverteilung II

Zufallsexperiment mit binärer Zufallsvariable



Binomialverteilung IV

Grundgesamtheit
Wahrscheinlichkeit für Erfolg p

↓ Stichprobe Umfang n

Eine Sequenz: (0, 1, 1, 0, 1, ..., 0, 1)

↓ unabhängige Experimente

Umsortierung

k Erfolge in n

$$\underbrace{(1, 1, 1, 1\ldots,}_{k}\ \underbrace{0, 0,\ldots, 0)}_{n-k}$$

Binomialverteilung V

- Wahrscheinlichkeit für genau eine Sequenz mit k Erfolgen und $n - k$ Misserfolgen:

$$p^k \cdot (1-p)^{n-k}$$

- folgt aus der vereinfachten Multiplikationsregel (V 4, F 31) für unabhängige Ereignisse

→ Baumdiagramm aus Vorlesung 4

Binomialverteilung VI

- Wie viele verschiedene Sequenzen mit *k*-Erfolgen gibt es?

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$ n über k → Binomialkoeffizient

$$n! = 1 \cdot 2 \cdot 3 \ldots \cdot n$$ n Fakultät

$$\binom{n}{0} = \binom{n}{n} = 1$$ weil 0! = 1 per Definition

z. B. Anzahl der Möglichkeiten aus 5 Buchstaben 2 auszuwählen:

$$\binom{5}{2} = \frac{5!}{2!\,3!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5}{1 \cdot 2 \cdot 1 \cdot 2 \cdot 3} = 10$$

Probe: AB, AC, AD, AE, BC, BD, BE, CD, CE, DE

Binomialverteilung VII

- Wahrscheinlichkeit, dass die Zufallsvariable X (d.h. Zahl der Erfolge) genau k Erfolge in n Versuchen registriert:

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

- Man sagt: X ist binomialverteilt, geschrieben:

$$X \sim Bi(n, p)$$

- Die Binomialverteilung hat als erste zentrale Momente (vgl. Folien 9 f)

Erwartungswert: $\mu = n \cdot p$

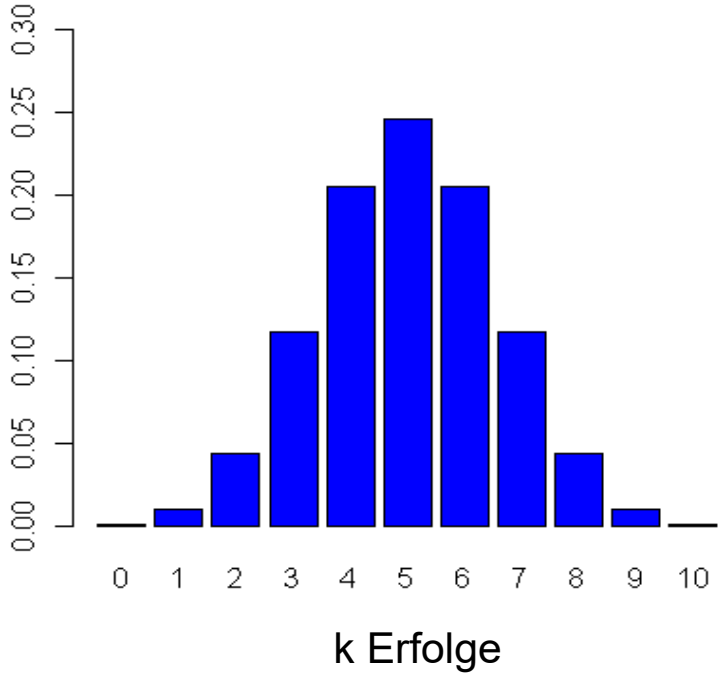
Varianz (der theoretischen Verteilung): $\sigma^2 = n \cdot p(1-p)$

Binomialverteilung VIII

(Stichprobe $n = 10$)

$p = 0{,}5$

$P(X = k)$



$\mu =$

$\sigma^2 =$

$\sigma =$



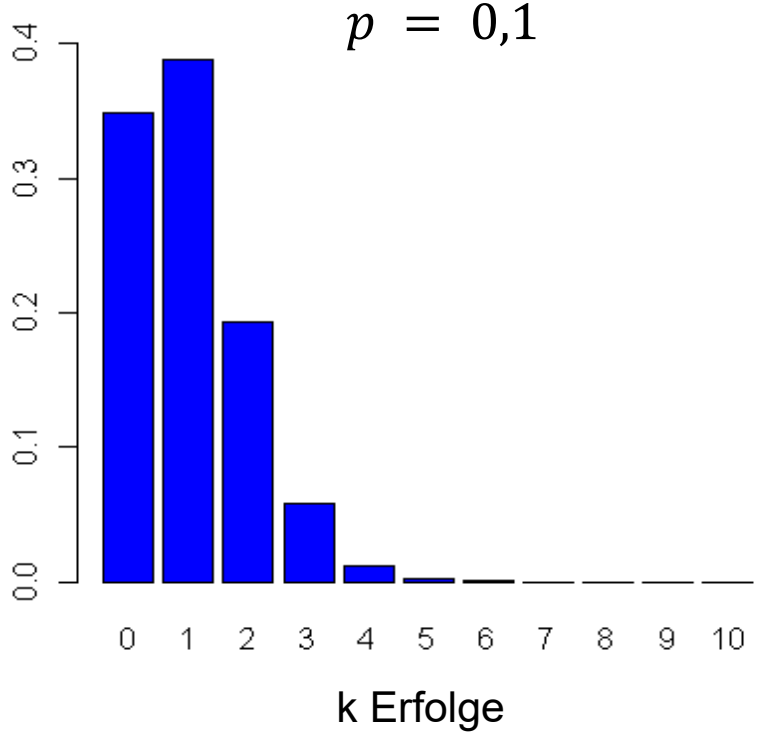
Gonick & Smith: Statistik in Cartoons

Binomialverteilung IX

(Stichprobe $n = 10$)

$p = 0{,}1$

$P(X = k)$



$\mu =$

$\sigma^2 =$

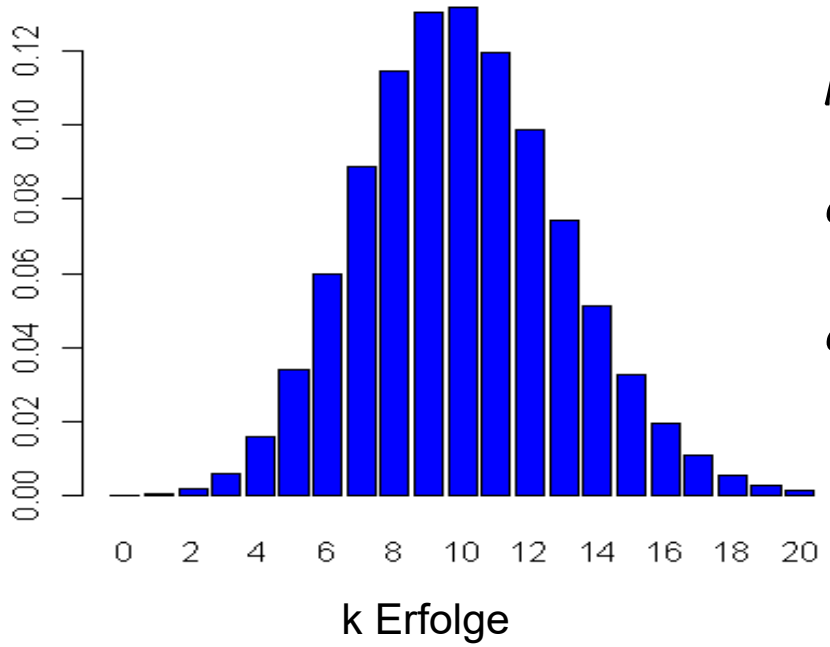
$\sigma =$

Binomialverteilung X

(Stichprobe $n = 100$)

$p = 0{,}1$

$P(X = k)$



$\mu =$

$\sigma^2 =$

$\sigma =$

Binomialverteilung – Beispiel B 5.3

30 % der Sprechstunden-Patienten mit einer Diagnose haben Fieber.
Wie groß ist die Wahrscheinlichkeit, dass von 10 Patienten genau zwei Fieber haben?

$p = 0{,}3; n = 10; k = 2$

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

$P(X = 2) =$

Warum könnte diese Rechnung nicht ganz aufgehen?



Jotquadrat: Grafitto in Bonn. commons. wikipedia.org

Zentraler Grenzwertsatz: Binomialverteilung

- Die Binomialverteilung wird durch die Normalverteilung approximiert (ZGWS von Moivre 1730 / Laplace 1812), wobei:

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)}$$

$$B(n,p) \rightarrow N(np, np(1-p))$$

- Die Approximation ist zufriedenstellend, falls gilt:

$$np \geq 5 \quad \textit{und} \quad n \cdot (1-p) \geq 5$$

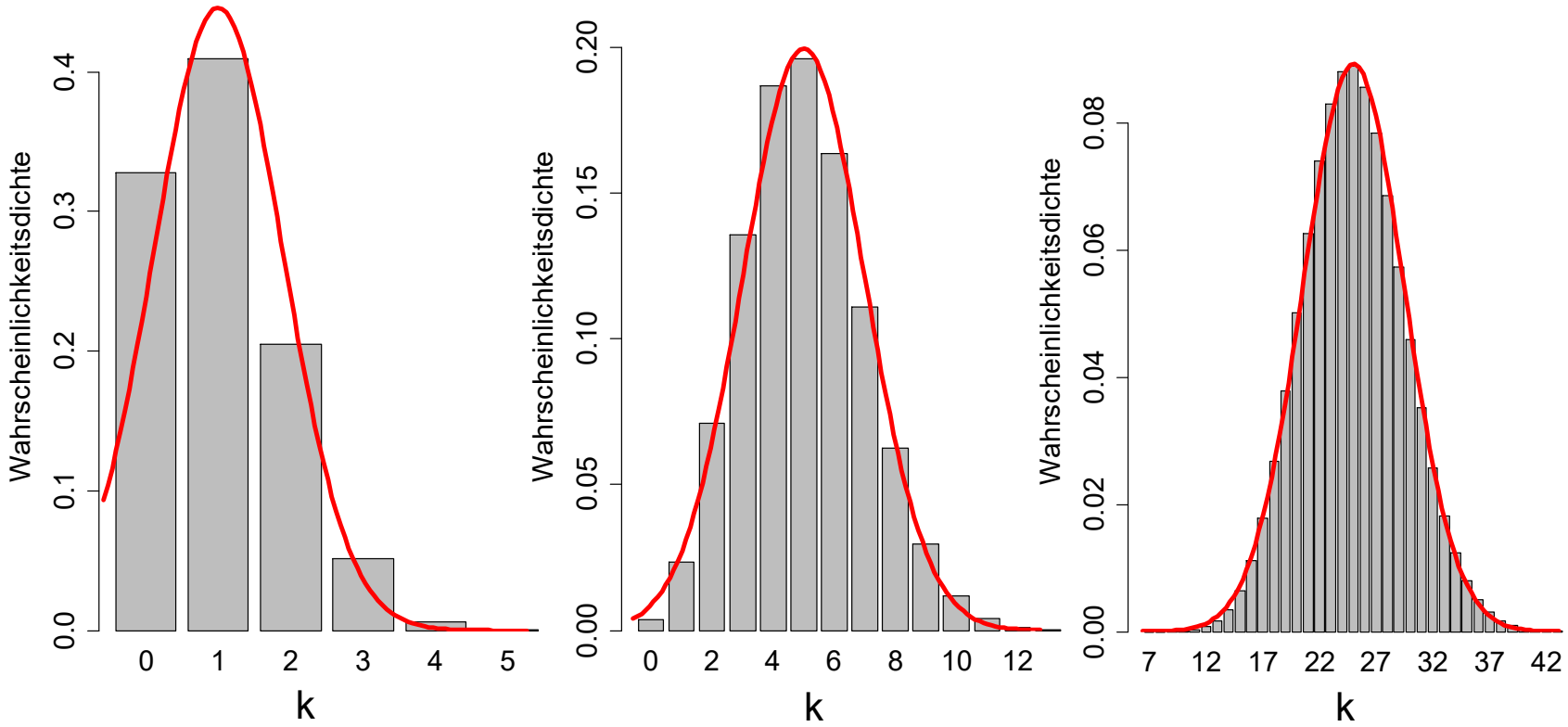


Gonick & Smith: Statistik in Cartoons

ZGWS: Binomialverteilung 2

- Für $n \to \infty$ geht die Binomial- in die Normalverteilung über:

$p = 0{,}2;\ n = 5$ $p = 0{,}2;\ n = 25$ $p = 0{,}2;\ n = 125$

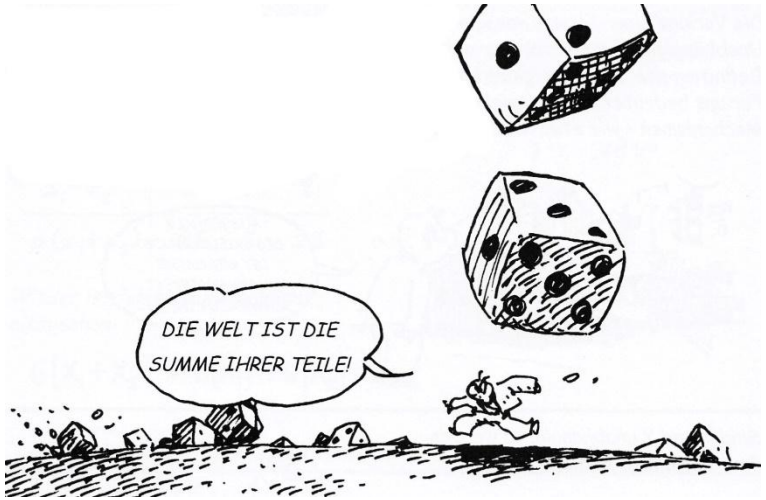


V 5

5.2 Normalverteilung

Modellierung mit Normalverteilung

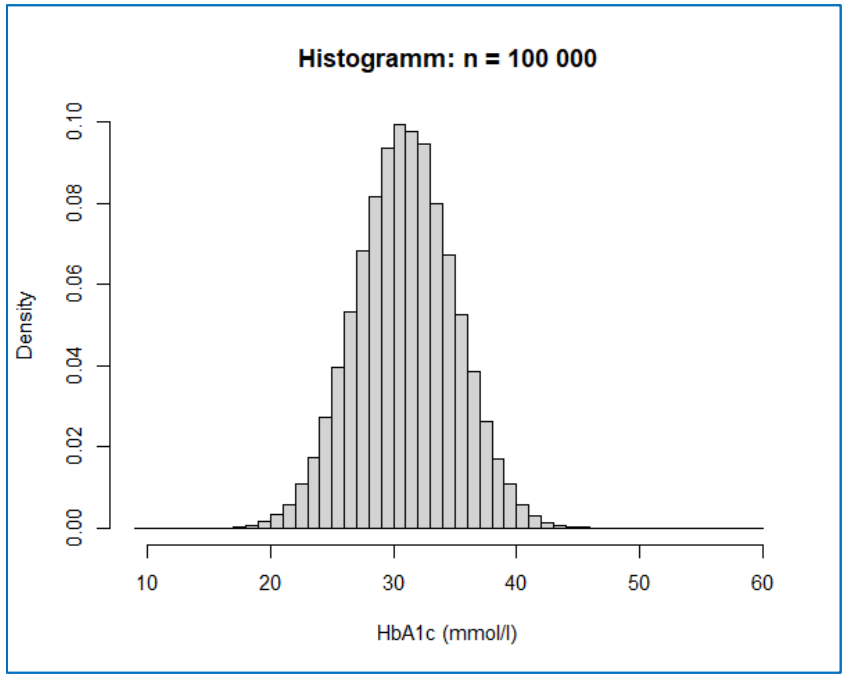
- Auf viele Prozesse in der Natur kann die Normalverteilung gut angewandt werden.
- Daten, die durch viele kleine und unabhängige Zufallseffekte beeinflusst werden, sind approximativ normalverteilt.
- Beispiel: mittlere Jahrestemperatur, HbA1c bei Gesunden (F. 31)

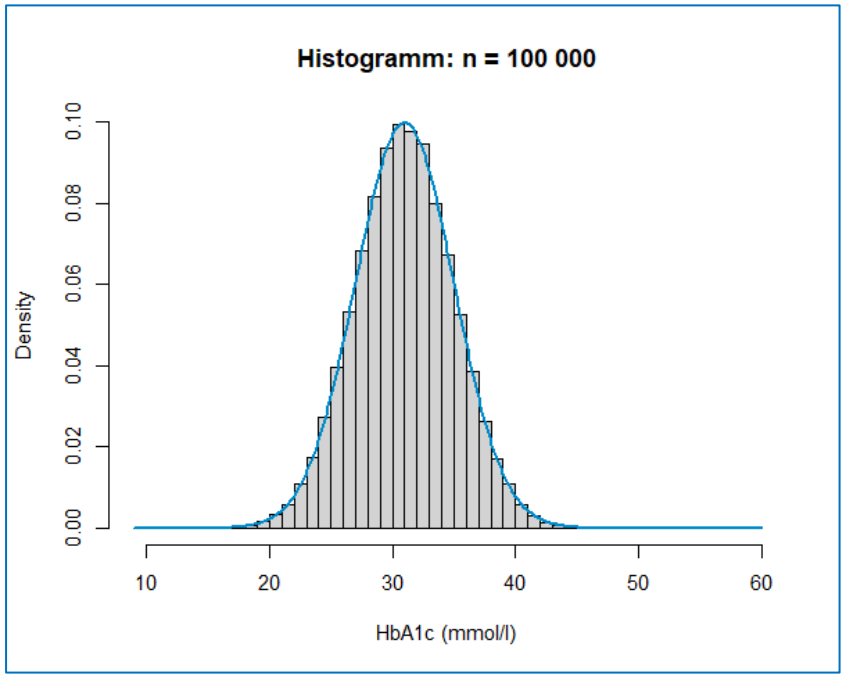


Gonick & Smith: Statistik in Cartoons

Dichtefunktion stetiger ZV

Erstellt man Histogramme eines Merkmals von immer größeren Stichproben, hier z. B. HbA1c, konvergiert die Umrisslinie gegen eine Kurve, die Dichtekurve der ZV.

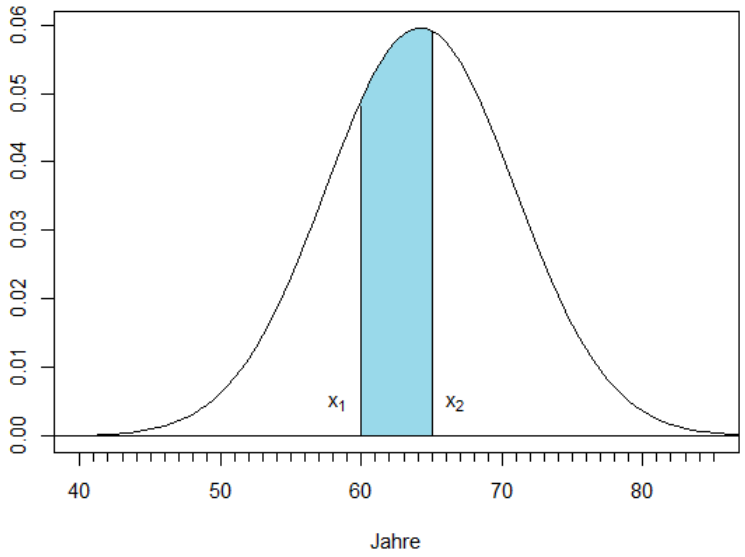




Normalverteilung I

Wahrscheinlichkeits-Interpretation:

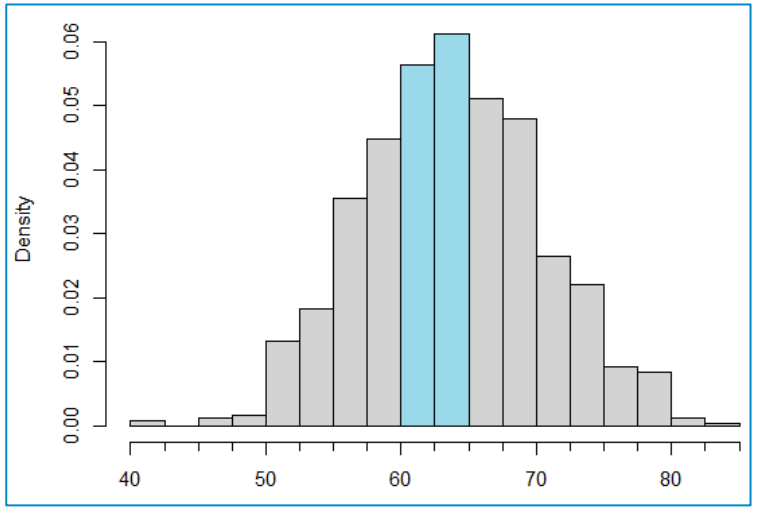
- Die Fläche unter einer Dichtefunktion zwischen zwei Fußpunkten x_A, x_B ist als Wahrscheinlichkeit (bei empirischen Kurven als relative Häufigkeit) interpretierbar.



$$F(-\infty, +\infty) = 1$$

$$F(x_1, x_2) < 1$$

$$F(x, x) = 0$$



Normalverteilung II

- Dichtefunktion der Normalverteilung:

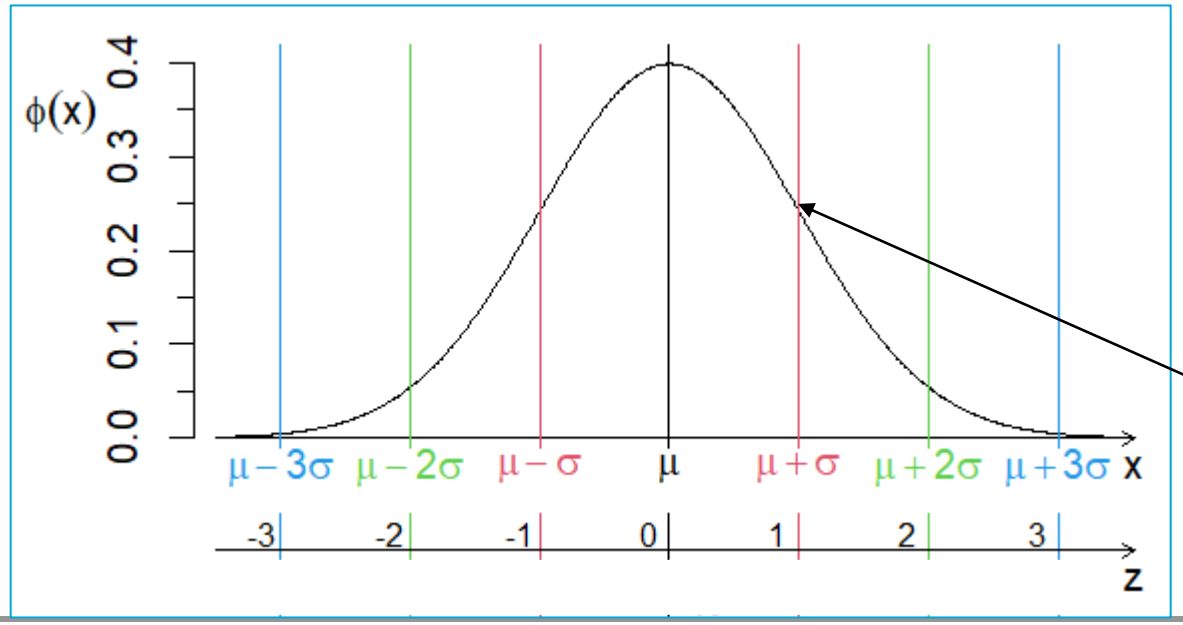
$$N(\mu, \sigma^2)$$

Lage-parameter

Streuungs-parameter

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

$$e = 2{,}718$$



Glockenkurve:

- Symmetrie
- Wendepunkte

Normalverteilung III

- Carl Friedrich Gauß (1777 – 1855)
- Mathematiker, Kartograph und Astronom
- Biographie: Daniel Kehlmann:
  Die Vermessung der Welt

- Gauß entdeckte die Bedeutung der Normalverteilung (z. B. für astronomische Betrachtungen) und wandte diese an.



Normalverteilung IV

- Dichtefunktion der Standardnormalverteilung:

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



- Verteilungsfunktion der Standardnormalverteilung:

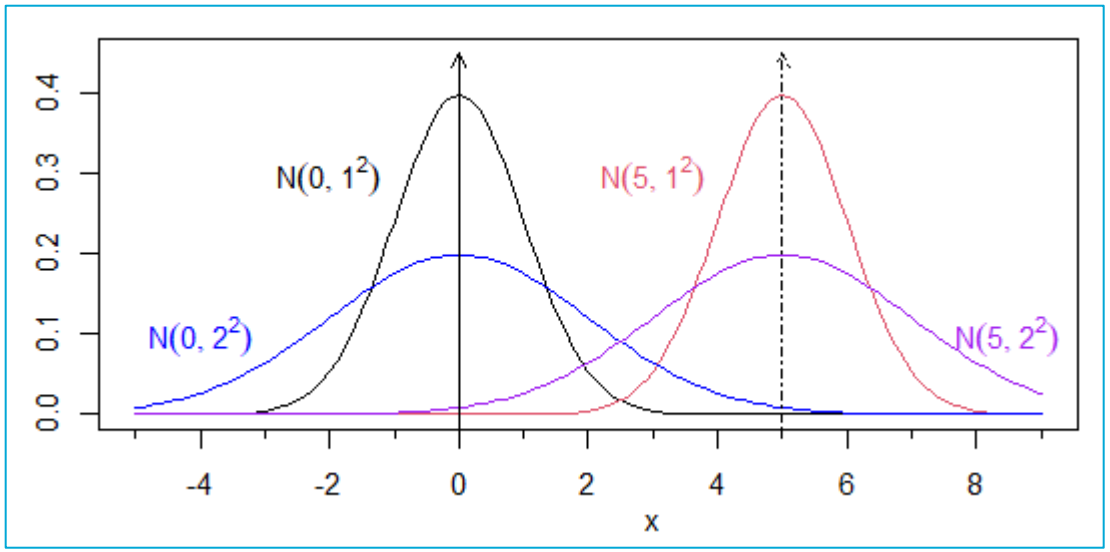
(gibt die Wahrscheinlichkeit an, dass eine stetige Zufallsgröße höchstens den Wert z annimmt)

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \mathrm{e}^{\frac{-y^2}{2}}\,\mathrm{dy}$$

Normalverteilung V

Die Normalverteilung ist eindeutig durch 2 Parameter festgelegt:

- Mittelwert μ: Maximum der Glockenkurve
- Varianz σ^2: Breite der Glockenkurve

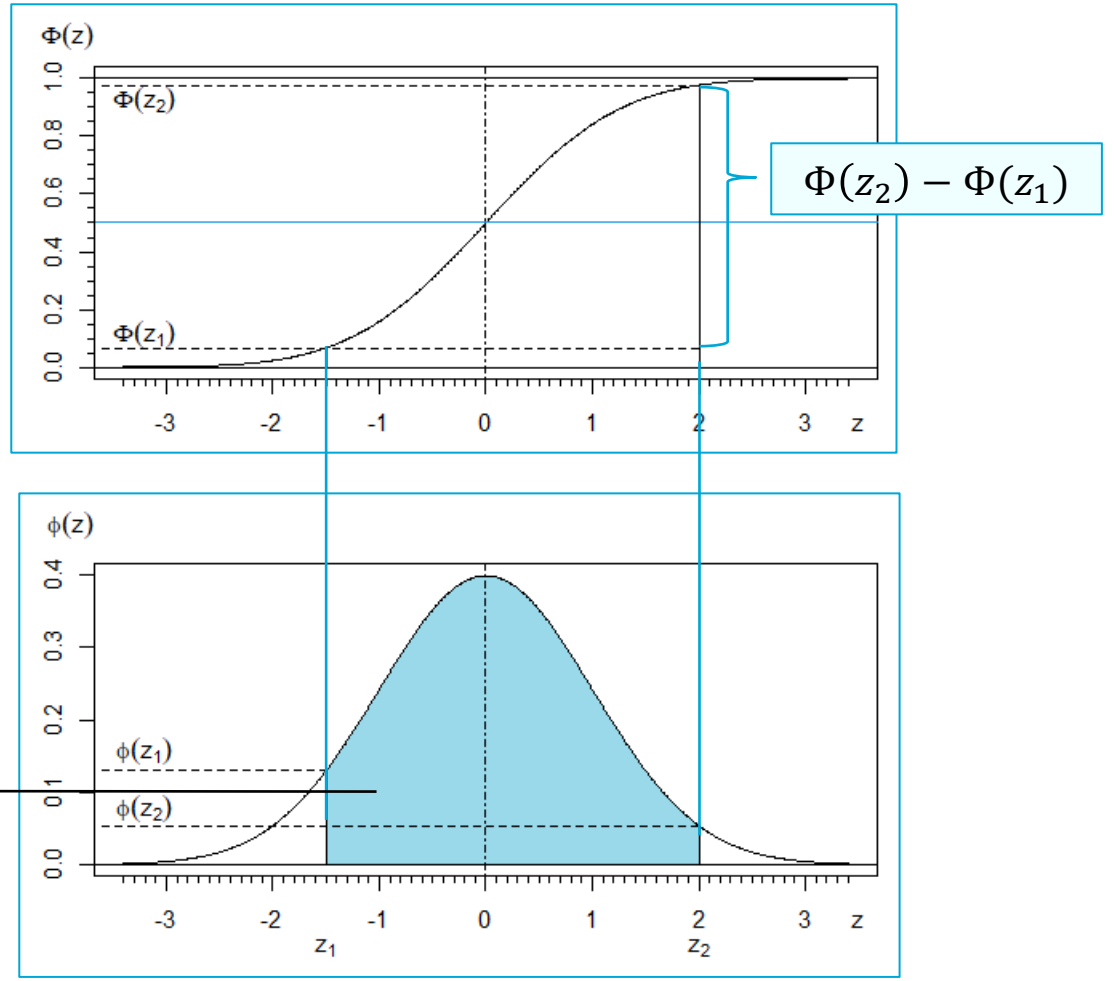


Normalverteilung VI

Verteilungsfunktion

- $\Phi(z) = \int_{-\infty}^{z} \varphi(x)dx$

- Dichtefunktion



$P(z_1 < Z < z_2)$

Standardisierung

- Idee: Standard-Normalverteilung N(0, 1) als Grundlage für alle Normalverteilungen

- Standardisierung*:* Durch Transformation (Skalenumrechnung) kann man jede Normalverteilung $N(\mu, \sigma^2)$ in eine Standard-Normalverteilung $N(0,1)$ überführen:

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad Z \sim N(0,1)$$

$$Z = \frac{X - \mu}{\sigma}$$ z-Transformation

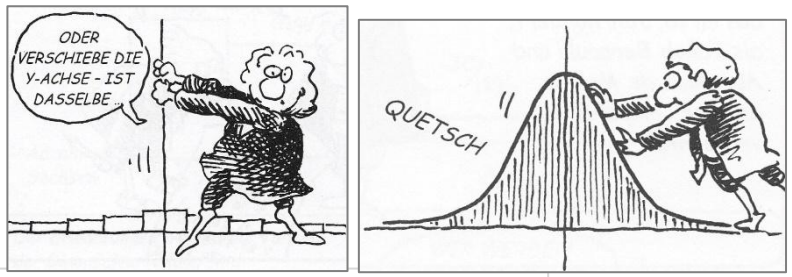


Gonick & Smith: Statistik in Cartoons

Transformation

- Umgekehrt lässt sich jede Normalverteilung $N(\mu, \sigma^2)$ durch lineare Transformation aus der Standard-Normalverteilung gewinnen:

$$X = Z \cdot \sigma + \mu$$



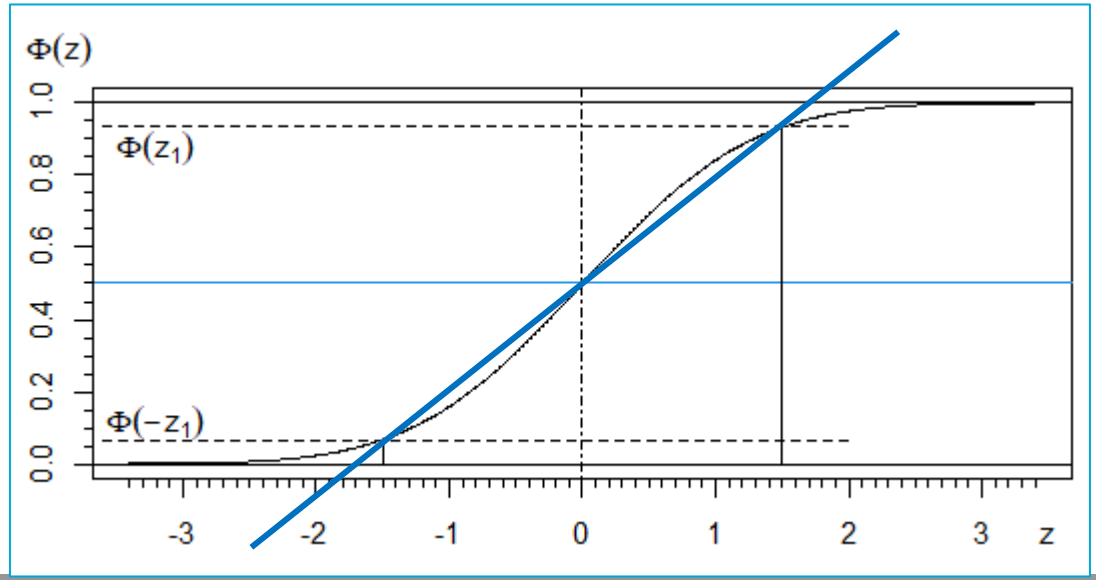
Gonick & Smith: Statistik in Cartoons

$$X \sim N(\mu, \sigma^2) \quad \Leftarrow \quad Z \sim N(0,1)$$

Tabellierung der Normalverteilung

Die Verteilungsfunktion $\Phi(z)$ der Standardnormalverteilung ist symmetrisch zum Erwartungswert 0. Daher müssen nur Werte für $z \geq 0$ tabelliert werden. Es gilt nämlich:

1. $\Phi(-z) = 1 - \Phi(z)$ oder $\Phi(-z) + \Phi(z) = 1,$
2. $\Phi(Z \geq z) = 1 - \Phi(Z < z)$



Wahrscheinlichkeiten ablesen

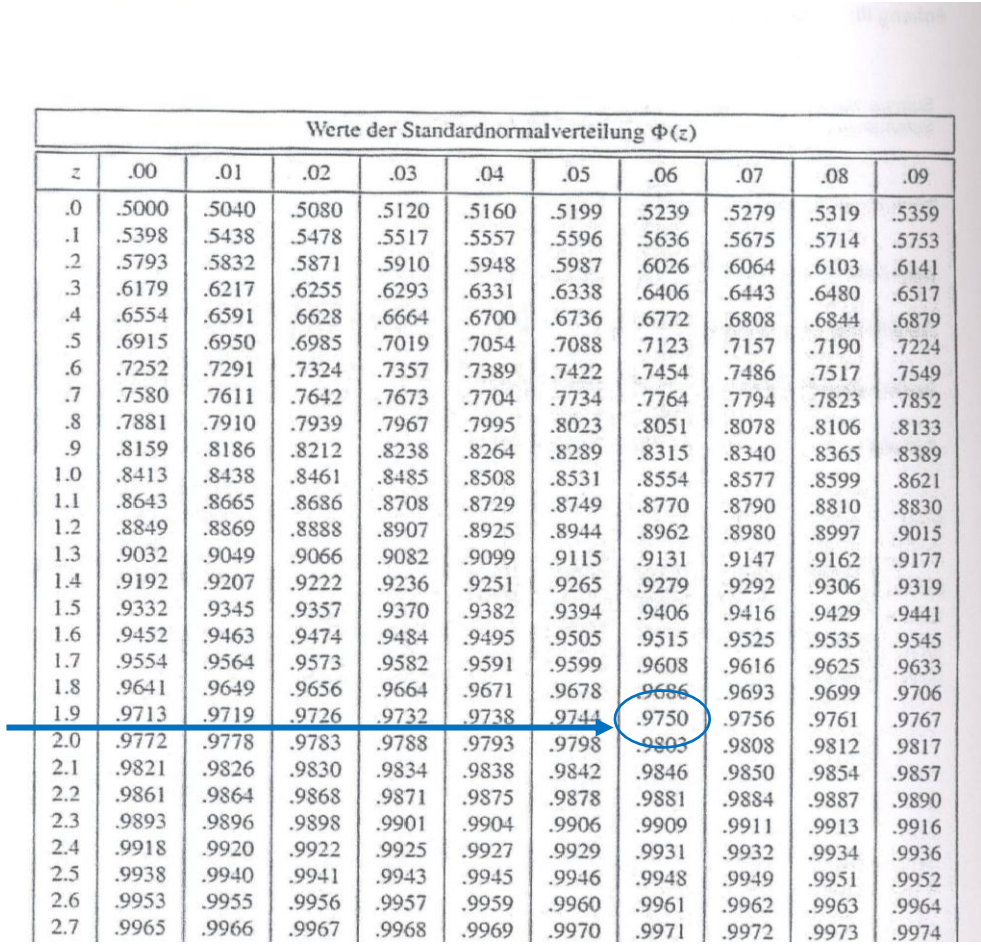
Verteilungsfunktion $\Phi(z)$ der Standardnormalverteilung:

- Skriptenheft S. 176, Tab. 11.1

$$\Phi(z = 1{,}96)$$
$$= 0{,}975 = 97{,}5\%$$

Umgekehrt ist $z_{0{,}975} = 1{,}96$

- das 97,5%-Quantil der Standardnormalverteilung



Werte der Standardnormalverteilung $\Phi(z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6338	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7252	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9416	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974

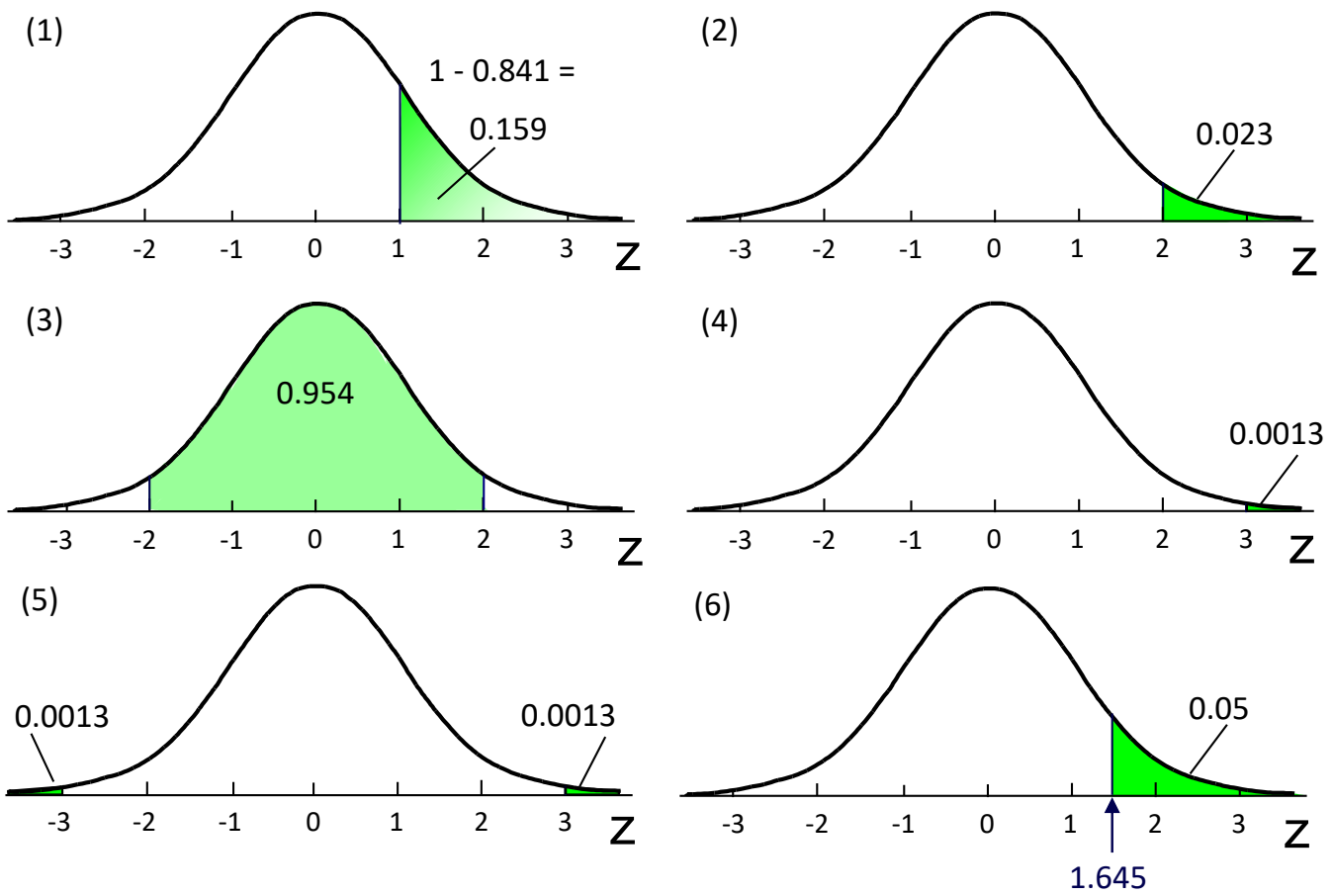
(aus: Guggenmoos-Holzmann, Irene u.a.: Medizinische Statistik. Blackwell Wissenschaftsverlag. Berlin, Wien 1996)

Quantile der Standard-NV ablesen

Quantile der standardisierten Normalverteilung N (0,1):

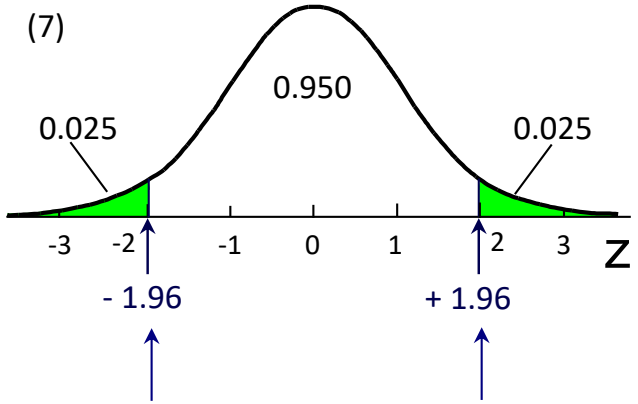
α		0,1000	0,0500	0,020	0,010	0,005
einseitig:	$z_{1-\alpha}$	1,2815	1,645	2,053	2,326	2,575
zweiseitig:	$z_{1-\alpha/2}$	1,6450	1,960	2,326	2,575	2,810

Quantile der Standardnormalverteilung



Quantil für das 95%-C.I.

- Konfidenzintervalle auf Basis der Normal- oder t-Verteilung sind symmetrisch. Deshalb braucht man dafür das $0{,}025$- bzw. $0{,}975$-Quantil, dessen Betrag $1{,}96 \approx 2$ ist. Eine meist ausreichende Näherung für das 95%-C.I. des Mittelwerts ist deshalb: $\bar{x} \pm 2 \cdot S.E.M.$

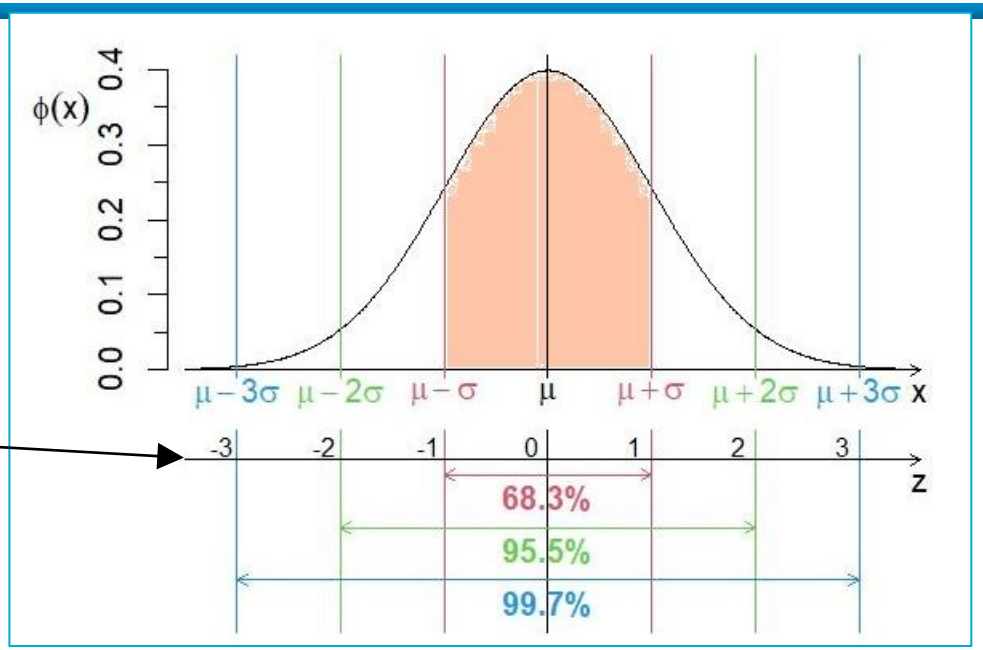


$$\alpha = 0{,}05$$
$$z_{1-\alpha/2} = 1{,}96$$

Anwendung: Norm- und Referenzbereiche

- basierend auf Normalverteilung

standardisierte (z-) Achse



- $\mu \pm 2\sigma$ Intervall zur Datenbeschreibung
- $\mu \pm 1{,}96\sigma$ Intervall als Referenzbereich (umfasst 95% der Verteilung)
- Auf dieser Basis werden häufig Referenzbereiche auf der Grundlage von Blutspenderdaten festgelegt.

Wahrscheinlichkeiten in einem Intervall

Gegeben seien

- $F(x)$ die Verteilungsfunktion einer Normalverteilung $N(\mu, \sigma^2)$,
- $\Phi(z)$ die Verteilungsfunktion der Standardnormalverteilung
- und $[a, b]$ ein Intervall $(b > a)$.

Dann berechnet sich die Wahrscheinlichkeit, dass ein Wert x dieser Normalverteilung zwischen a und b liegt:

$$P(a \leq x \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Ablauf:

- Standardisieren,
- Tafelwerte ablesen und
- Differenz bilden

Normalverteilung - Beispiel II

Beispiel 5.4: HbA_{1C}-Werte ~ normalverteilt $N(5{,}34\%, 0{,}18\ \%)$

- Der obere Referenzwert wird mit 5,7 % angegeben.
- Wie groß ist der Anteil der Gesunden mit einem $HbA_{1C} \leq 5{,}7\%$?

Ablauf:

- Standardisierung: $z =$
- Tafelwert: Skriptenheft S. 176, Tabelle 11.1: $\Phi(z) =$
- also folgt:

Tabelle 11.1: *Verteilungsfur*

z	.00	.01
.0	.5000	.5040
.1	.5398	.5438
.2	.5793	.5832
.3	.6179	.6217
.4	.6554	.6591
.5	.6915	.6950
.6	.7252	.7291
.7	.7580	.7611
.8	.7881	.7910
.9	.8159	.8186
1.0	.8413	.8438
1.1	.8643	.8665
1.2	.8849	.8869
1.3	.9032	.9049
1.4	.9192	.9207
1.5	.9332	.9345
1.6	.9452	.9463
1.7	.9554	.9564
1.8	.9641	.9649
1.9	.9713	.9719
2.0	.9772	.9778

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

Vielen Dank!

Dr. rer. nat. Meinhard Mende

Senior-Wissenschaftler, Institut für Medizinische Informatik, Statistik und Epidemiologie

Härtelstraße 16-18, 04107 Leipzig

Meinhard.mende@zks.uni-leipzig.de



© MM