

M24 Statistik 1: Sommersemester 2024

Vorlesung 09: Wahrscheinlichkeitsdichte

Prof. Matthias Guggenmos

Health and Medical University Potsdam

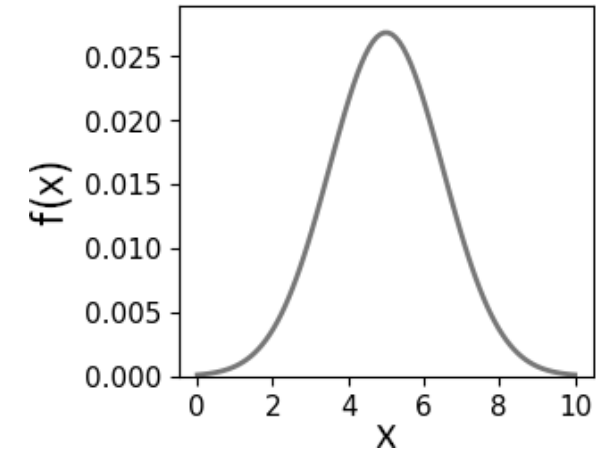


Theoretische Häufigkeitsverteilungen

In der letzten Vorlesung haben wir erarbeitet, dass sich die Stichprobenverteilung durch eine Normalverteilung der Form

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

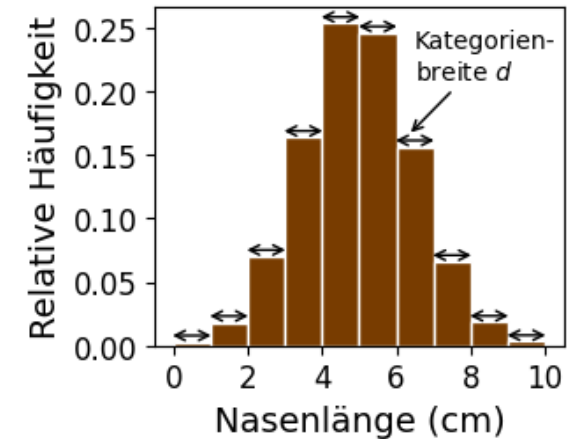
beschreiben lässt (wobei die Variable x im Fall der Stichprobenverteilung der Stichprobenkennwert $\hat{\theta}$ ist).



- Eine theoretische Häufigkeitsverteilung wie die Normalverteilung gibt für *jeden beliebigen Wert* x des Merkmals X eine Häufigkeit $f(x)$ an.
- Eine wichtige Frage haben wir bislang jedoch nicht beantwortet: was für eine Art von Häufigkeit ist $f(x)$ an? Wie ist also die y-Achse im Diagramm rechts oben zu interpretieren?

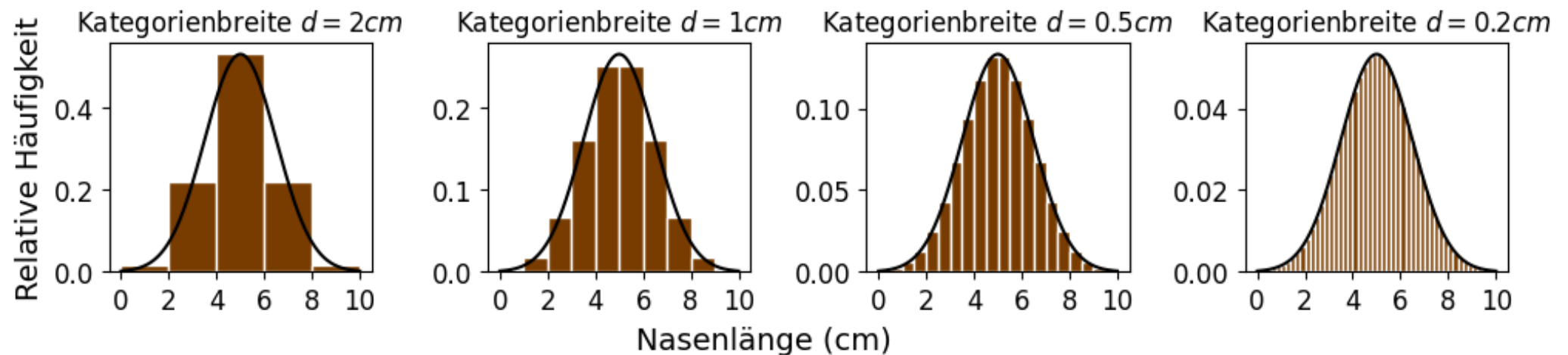
Rückblick: das Histogramm

- Um die Bedeutung von $f(x)$ zu verstehen, gehen wir zunächst zurück zum **Histogramm**.
- Histogramme stellen die Häufigkeit der Merkmalsvariable X in einer Stichprobe oder Population dar. Wird die **relative Häufigkeit** aufgetragen, so ordnet das Histogramm jedem Intervall auf der x-Achse eine relative Häufigkeit dar, die z.B. in Prozent angegeben werden kann.
- Die Breite des Intervalls – die **Kategorienbreite d** – ist dabei ein Kompromiss zweier Faktoren:
 1. **Auflösung:** je schmaler das Intervall, desto feiner wird das Merkmal X unterteilt.
 2. **Fallzahl:** je breiter das Intervall, desto höher die Zahl der Fälle im Intervall, desto präziser die Schätzung des Häufigkeitswertes im Intervall.
- Die Gesamtsumme aller Säulen im Histogramm mit relativer Häufigkeit ist immer 1 (oder 100%).



Rückblick: das Histogramm

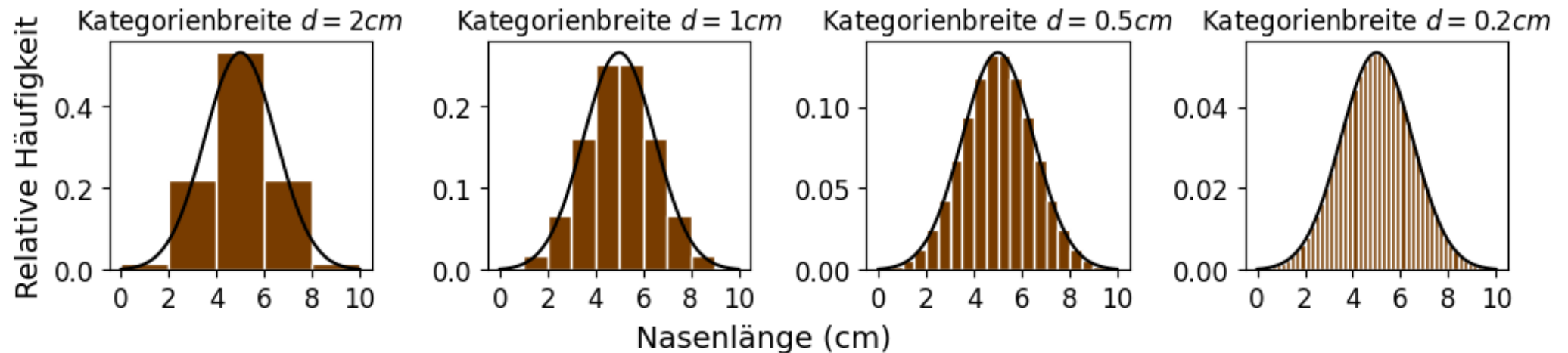
- Wir nehmen nun an, dass wir das Histogramm auf Basis einer **unendlich großen** Population bilden, in der die theoretische Häufigkeitsverteilung von Nasenlängen durch eine **Normalverteilung** beschrieben wird (im Beispiel: $\mu = 5\text{cm}$, $\sigma = 1.5\text{cm}$).
- Egal, wie fein wir die Kategorienbreite wählen, gibt es aufgrund der unendlich großen Population genug Datenpunkte für eine präzise Schätzung der relativen Häufigkeit in einer Säule.
- Es gilt: je kleiner die Kategorienbreite, desto mehr Säulen gibt es, desto kleiner die relativen Häufigkeitswerte jeder einzelnen Säule.



- Es zeigt sich: mit feiner werdender Kategorienbreite nähert sich das Histogramm einer Normalverteilung — und damit der theoretische Häufigkeitsverteilung — an!
(→ Gesetz der großen Zahlen)

Rückblick: das Histogramm

- Nun scheint ein Brückenschlag naheliegend: ist die theoretische Häufigkeitsverteilung $f(x)$, die Funktionswerte für beliebige x -Werte ausgibt, gleich einem Histogramm, bei dem die Kategorienbreite gegen Null geht?



- Im Prinzip ja, allerdings gibt es noch ein Problem: geht die Kategorienbreite d gegen 0, gehen die relativen Häufigkeitswerte des Histogramms ebenfalls gegen Null!
- Würde die theoretische Häufigkeitsverteilung $f(x)$ also relative Häufigkeiten angeben, so wäre $f(x)$ für jedes x Null. Das ist natürlich sinnlos.
- **Theoretische Häufigkeitsverteilungen $f(x)$ geben aus diesem Grund keine relative Häufigkeit an.**
- Bleibt die Frage: was stattdessen?

Von der Wahrscheinlichkeit zur Wahrscheinlichkeitsdichte

Zunächst ein Hinweis zur Nomenklatur:

Defin ition	Wir verwenden den Begriff relative Häufigkeiten bei empirischen Daten und meinen damit den Anteil einer Merkmalsausprägung relativ zu allen Datenpunkten. Beispiel: in einer Stichprobe von 100 Würfelversuchen lag die relative Häufigkeit von Zahlen größer 3 bei 0.48 oder 48%.
------------------------	---

Defin ition	Wir verwenden den Begriff Wahrscheinlichkeit , wenn die theoretische Häufigkeitsverteilung eines Merkmals bekannt ist, und meinen damit den Anteil einer Merkmalsausprägung laut Theorie. Beispiel: bei einem perfekten Würfel ist die Wahrscheinlichkeit einer Zahl größer 3 exakt 0.5.
------------------------	---

- Im Kontext von theoretischen Häufigkeitsverteilungen können wir daher von **Wahrscheinlichkeiten** sprechen.
- Klar ist auch: durch die Nomenklaturänderung *relative Häufigkeit* → *Wahrscheinlichkeit* ist noch nichts gewonnen.

Von der Wahrscheinlichkeit zur Wahrscheinlichkeitsdichte

Der entscheidende Trick theoretischer Häufigkeitsverteilungen ist der **Übergang von Wahrscheinlichkeiten zu Wahrscheinlichkeitsdichten**.

Ist das Merkmal X eine kontinuierliche Variable (z.B. Nasenlänge in cm), so geben theoretische Häufigkeitsverteilungen $f(x)$ eine **Wahrscheinlichkeitsdichte** an.

Wie kann man sich “Wahrscheinlichkeitsdichte” vorstellen?

- Wir kennen das Konzept der “Dichte” bei Stoffen: z.B. ist die Dichte von Eis ist ca. 1 g/cm^3 , d.h. dass sich eine Masse von $1g$ in einem Kubikzentimeter ($1cm^3$) befindet.
- Eine Dichte ist also immer eine bestimmte Masse *pro* Maßeinheit.

Von der Wahrscheinlichkeit zur Wahrscheinlichkeitsdichte

Wir können daher Wahrscheinlichkeitsdichte wie folgt definieren:

Defin
ition Wahrscheinlichkeitsdichte = Wahrscheinlichkeits(masse) *pro* Maßeinheit

- In Abgrenzung zur Wahrscheinlichkeits*dichte* wird die Wahrscheinlichkeit selbst tatsächlich auch als Wahrscheinlichkeits*masse* bezeichnet (engl. *probability mass*).
 - Jedoch ist Wahrscheinlichkeit bzw. Wahrscheinlichkeitsmasse im Gegensatz zur physikalischen Masse einheitslos.
- Die Einheit der Wahrscheinlichkeitsdichte wiederum ist Wahrscheinlichkeit *pro* Maßeinheit: Wahrscheinlichkeit *pro* Zentimeter Nasenlänge, Wahrscheinlichkeit *pro* IQ-Punkt, Wahrscheinlichkeit *pro* Fragebogenpunkt.

Wahrscheinlichkeitsdichte

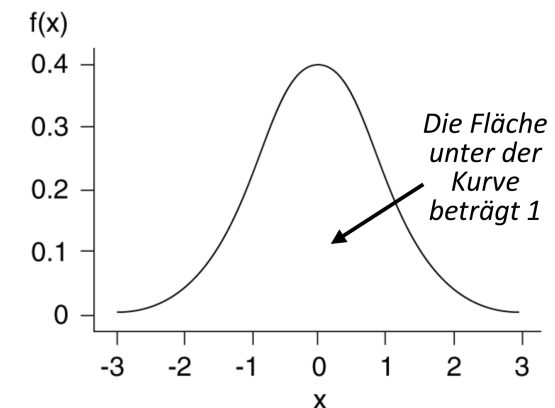
- Theoretische Häufigkeitsverteilungen $f(x)$ für kontinuierliche Merkmale X werden auch als **Wahrscheinlichkeitsdichtefunktion** bezeichnet (engl. *probability density function*).
- Wie bei Histogrammen mit relativen Häufigkeiten ist die *gesamte Wahrscheinlichkeitsmasse* von Wahrscheinlichkeitsdichtefunktionen $f(x)$ immer 1.
- Anders gesagt: Wahrscheinlichkeitsdichtefunktion $f(x)$ sind immer so normalisiert, dass ihr Flächeninhalt den Wert 1 hat.

Beispiel Normalverteilung:

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

Der Normalisierungsfaktor $\frac{1}{\sigma\sqrt{2\pi}}$ sorgt in diesem Fall dafür, dass die Fläche unter der Normalverteilung gleich 1 ist:

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

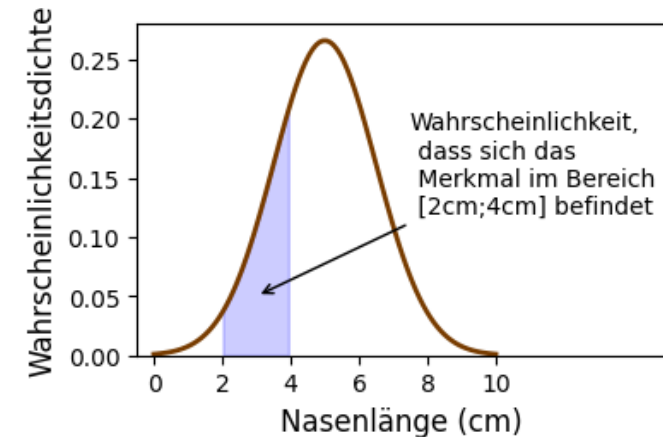


Von der Wahrscheinlichkeitsdichte zurück zur Wahrscheinlichkeit

- Um aus einer Wahrscheinlichkeits*dichte* eine Wahrscheinlichkeit zu erhalten, muss die Dichte über einen bestimmten **Wertebereich** $[x_0; x_1]$ des Merkmals summiert (integriert) werden.
- Mathematisch beschreiben wir diese Operation als ein Integral:

$$P(x_0 < x < x_1) = \int_{x_0}^{x_1} f(x) dx$$

- P ist die Wahrscheinlichkeit, dass das Merkmal einen Wert zwischen x_0 (Untergrenze) und x_1 (Obergrenze) aufweist.
- Das Integral setzt die *Wahrscheinlichkeitsdichte* $f(x)$ mit der *Wahrscheinlichkeit* $P(x_0 < x < x_1)$ in Verbindung.

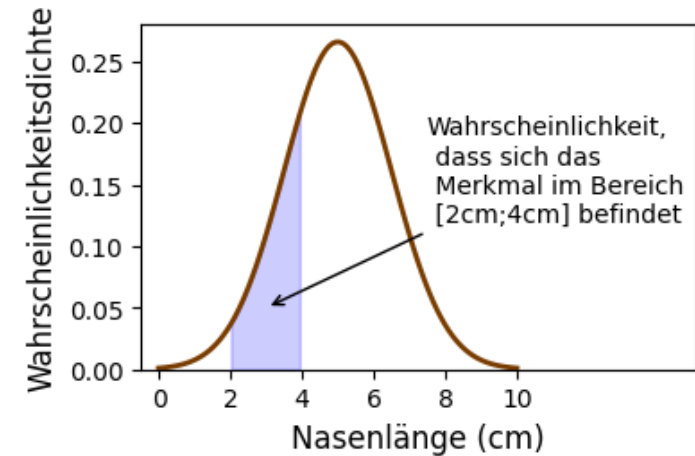


Berechnung einer Wahrscheinlichkeit P auf Basis einer Wahrscheinlichkeitsdichtefunktion $f(x)$ (hier der Normalverteilung).

Wahrscheinlichkeitsdichte: Beispiel 1

Nehmen wir an, dass Nasenlängen in der Population *normalverteilt* sind, mit Mittelwert $\mu = 5$ und Standardabweichung $\sigma = 1.5$.

Frage: wie hoch ist die Wahrscheinlichkeit, dass eine zufällig gezogene Nase aus der Population eine Länge zwischen 2cm und 4cm hat?

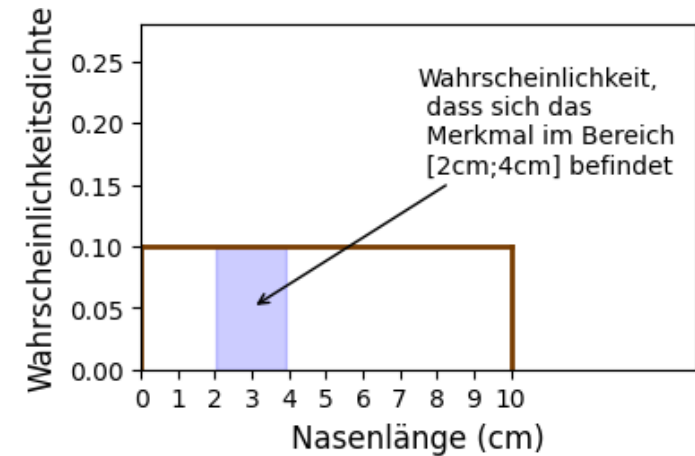


$$\begin{aligned}
 P(2 \leq x \leq 4) &= \int_2^4 f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_2^4 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \\
 &= \frac{1}{1.5\sqrt{2\pi}} \int_2^4 \exp\left(-\frac{(x-5)^2}{2 \cdot 1.5^2}\right) dx \stackrel{(Computer!)}{\approx} 0.23
 \end{aligned}$$

Wahrscheinlichkeitsdichte: Beispiel 2

Nehmen wir nun an, dass Nasenlängen in der Population *uniform* zwischen 0 und 10 cm verteilt sind.

Gleiche Frage: wie hoch ist die Wahrscheinlichkeit, dass eine zufällig gezogene Nase aus der Population eine Länge zwischen 2cm und 4cm hat?



Wir wissen: die Fläche unter der Verteilung muss 1 sein. Daher muss die Wahrscheinlichkeitsdichte für jeden Wert zwischen 0cm und 10cm gleich 0.1cm^{-1} betragen ($10\text{cm} \cdot 0.1\text{cm}^{-1} = 1$).

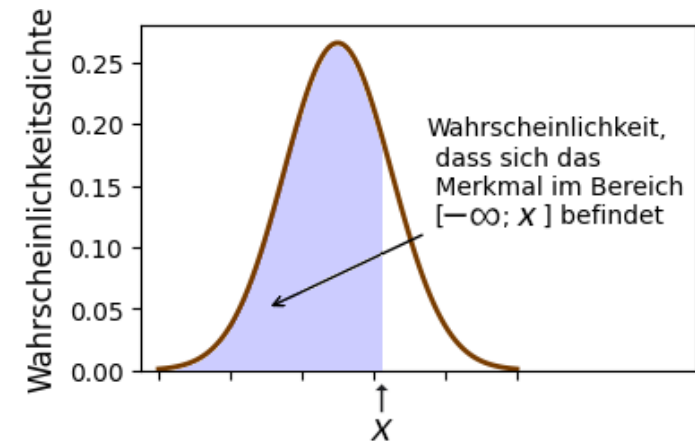
Die Berechnung des Flächeninhalts im Intervall $[2\text{cm}; 4\text{cm}]$ geht in diesem Fall ohne Integration, denn er entspricht einfach der Fläche eines Rechteckes mit Breite 2cm und Höhe 0.1cm^{-1} . Es gilt:

$$\begin{aligned} \text{Wahrscheinlichkeit} &= \text{Intervallbreite} \cdot \text{Wahrscheinlichkeitsdichte} = \\ &= 2\text{cm} \cdot 0.1\text{cm}^{-1} = 0.2 \end{aligned}$$

Verteilungsfunktion

Die Integration einer Wahrscheinlichkeitsdichte *bis zu einem bestimmten Wert x* ist ein sehr häufiger Fall im Umgang mit Wahrscheinlichkeitsdichten. Daher definieren wir dafür eine eigene Funktion, die **Verteilungsfunktion** $F(x)$:

$$F(x) = \int_{-\infty}^x f(x') dx'$$

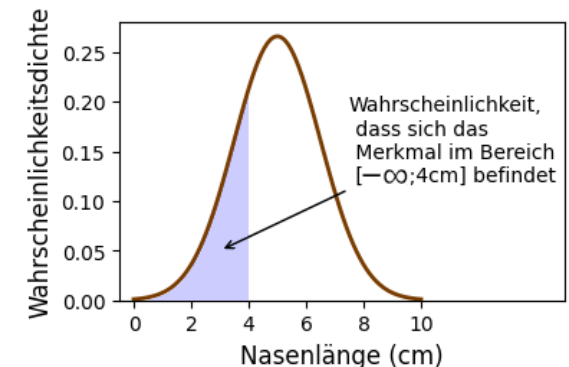


Die Verteilungsfunktion F gibt uns den Flächeninhalt der Dichtefunktion f “links von x ” an.

Nehmen wir wieder die normalverteilte Nasenlängen-Population an mit Mittelwert $\mu = 5$ und Standardabweichung $\sigma = 1,5$. Die Wahrscheinlichkeit, dass eine zufällig gezogene Nase eine Länge kleiner 4cm hat, ist gegeben durch den Wert $F(4)$ der Verteilungsfunktion dieser Normalverteilung:

Beispiel

$$\begin{aligned} F(4) &= \int_{-\infty}^4 f(x') dx' = \\ &= \frac{1}{1,5\sqrt{2\pi}} \int_{-\infty}^4 \exp\left(-\frac{(x' - 5)^2}{2 \cdot 1,5^2}\right) dx' \stackrel{(\text{Computer!})}{\approx} 0,25 \end{aligned}$$



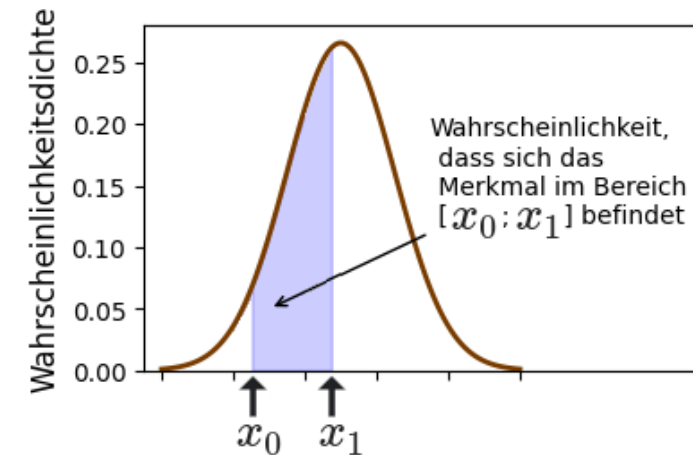
Verteilungsfunktion

Mithilfe der Verteilungsfunktion, lässt sich nun das Integral

$$P(x_0 < x < x_1) = \int_{x_0}^{x_1} f(x) dx$$

mit dem wir die Fläche zwischen einer Untergrenze x_0 und Obergrenze x_1 berechnen, auch folgendermaßen aufstellen:

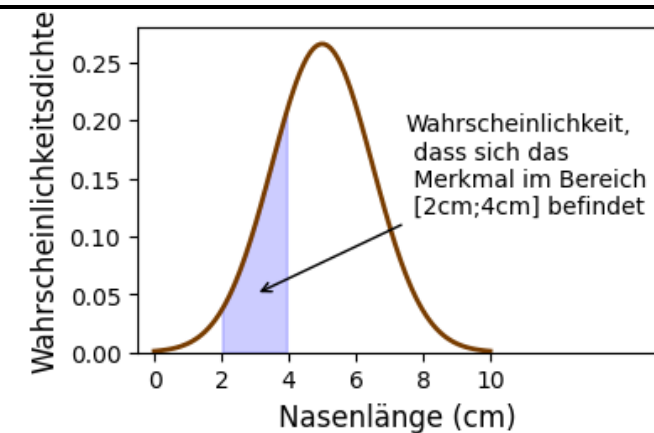
$$P(x_0 < x < x_1) = F(x_1) - F(x_0)$$



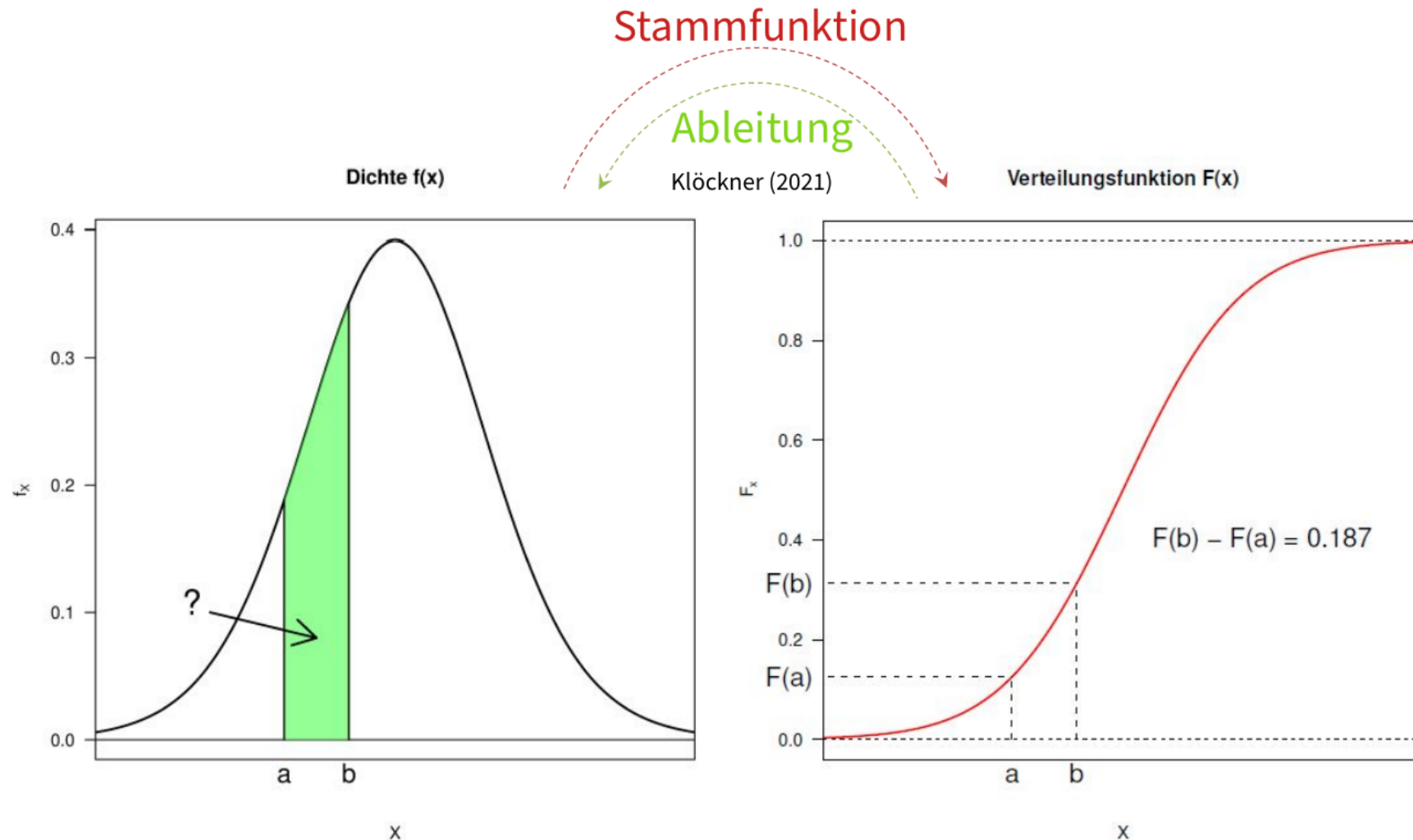
Beispiel

Die eingezeichnete Fläche aus unserem vorherigen Beispiel lässt sich berechnen als:

$$P(2 < x < 4) = F(4) - F(2) \stackrel{(Computer!)}{\approx} 0,23$$



Verteilungsfunktion und Stammfunktion

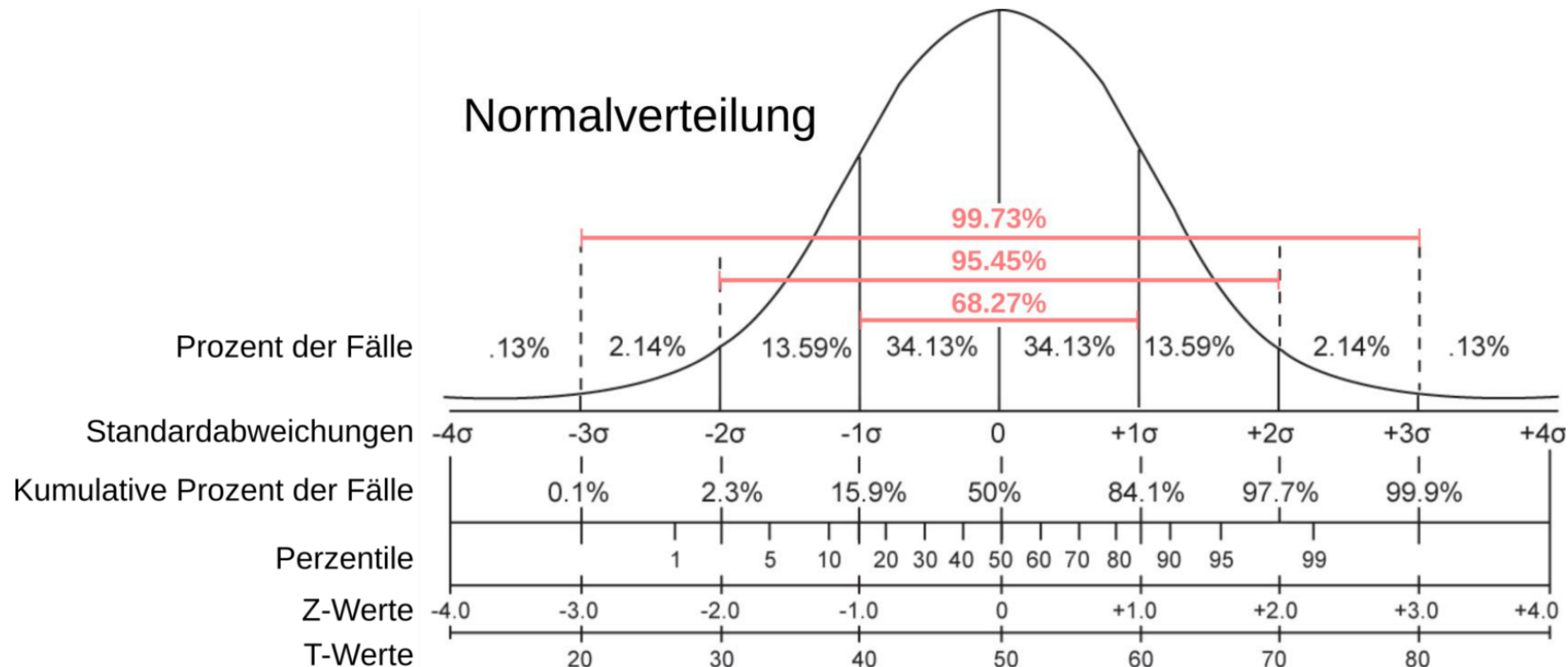


- $F(x)$ ist eine Stammfunktion von $f(x)$ wenn gilt: $\frac{dF}{dx} = f(x)$ bzw. $F(x) = \int_a^x f(x')dx'$.
- Die Verteilungsfunktion $F(x)$ entspricht der Stammfunktion $\int_a^x f(x')dx'$ mit $a = -\infty$.

68-95-99.7-Prozentregel

Mithilfe der Verteilungsfunktion lassen sich charakteristische Flächeninhalte der Normalverteilung berechnen. Als Faustregel ergibt sich die **68-95-99.7-Prozentregel**:

- Der Bereich Mittelwert \pm eine Standardabweichung ($\mu \pm 1\sigma$) umfasst **68%** der Daten
- Der Bereich Mittelwert \pm zwei Standardabweichungen ($\mu \pm 2\sigma$) umfasst **95%** der Daten
- Der Bereich Mittelwert \pm drei Standardabweichungen ($\mu \pm 3\sigma$) umfasst **99.7%** der Daten



Standardnormalverteilung

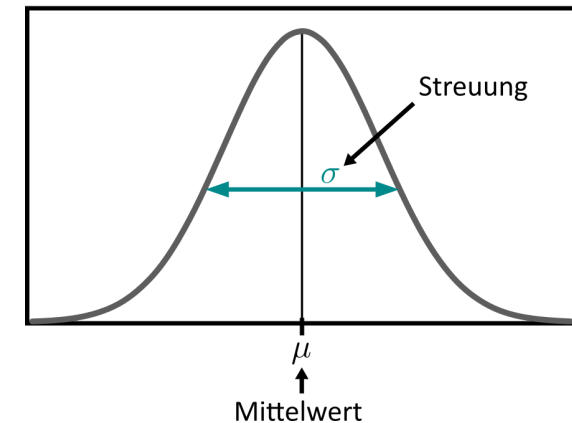
- Die Normalverteilung ist durch zwei Parameter charakterisiert, die Mittelwert und Standardabweichung der Verteilung definieren (idR μ und σ):

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

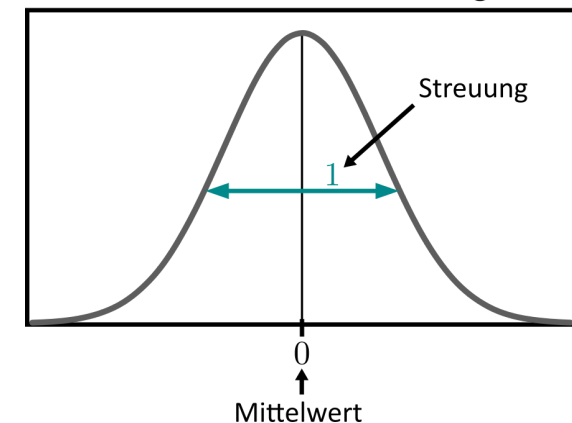
- Im weiteren Verlauf von Statistik 1 werden wir häufig die standardisierte Form der Normalverteilung verwenden – die **Standardnormalverteilung**.
- Die Standardnormalverteilung hat Mittelwert 0 und Standardabweichung 1:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \begin{matrix} \mu=0 \\ \sigma=1 \end{matrix} \quad \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Unstandardisierte Normalverteilung



Standardnormalverteilung



Das ABC der Normalverteilung

Aufgrund ihrer Bedeutung in der Statistik, haben sich für die (Standard)Normalverteilung bestimmte Bezeichnung eingebürgert, auf die wir ab jetzt zugreifen werden.

Normalverteilung mit Angabe des Mittelwertes μ und der Varianz σ^2	$\mathcal{N}(\mu, \sigma^2)$
Standardnormalverteilung (Mittelwert 0, Varianz 1)	$\mathcal{N}(0, 1)$
Dichtefunktion der Normalverteilung	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Dichtefunktion der Standardnormalverteilung	$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ (sprich “Klein Phi”) Es gilt: $f(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right)$
Verteilungsfunktion der Normalverteilung	$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{x'-\mu}{\sigma}\right)^2} dx$
Verteilungsfunktion der Standardnormalverteilung	$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}x'^2} dx$ (sprich “Groß Phi”) Es gilt: $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Vorschau

Im nächsten Schritt kehren wir zurück zur **theoretischen Stichprobenverteilung**. Die Erkenntnisse zur Wahrscheinlichkeitsdichte und Verteilungsfunktion lassen sich auf die theoretische Stichprobenverteilung übertragen und eröffnen so zwei wesentliche Methoden der Inferenzstatistik:

- **Hypothesentestung** bzw. **Signifikanztestung** (u.a. auch Idee des p-Wertes)
- **Konfidenzintervalle** (Verallgemeinerung des Standardfehlers)

