Mathematische und statistische Methoden für Pharmazeut*innen

Prof. Dr. Noemi Kurt
FB 12, Institut für Mathematik, Goethe-Universität Frankfurt

Sommersemester 2023

Vorlesung 10

Inhalt

- Eine Anwendung der Diffusionsgleichung
- Beschreibende Statistik: Daten und ihre Darstellungen
- Kenngrößen von Daten
- Häufigkeiten und Wahrscheinlichkeiten
- Berechnung von elementaren Wahrscheinlichkeiten

Lernziele

- Ein Anwendungsbeispiel der Diffusionsgleichung kennen
- Verschiedene Arten von Daten und Möglichkeiten für ihre Darstellung kennen
- Empirisches Mittel, Median, empirische Varianz und empirische Perzentile berechnen und interpretieren können
- Elementare Wahrscheinlichkeiten berechnen können

Benötigte Vorkenntnisse

- Funktionen, Ableitung, Ableitungsregeln; Zahlenmengen

Ein Beispiel für partielle Differentialgleichungen

Partielle Differentialgleichungen enthalten partielle Ableitungen nach verschiedenen Variablen.

Beispiel: Diffusion (erstes Fick'sches Gesetz)

$$\frac{\partial m}{\partial t} = -DA\frac{\partial c}{\partial x},$$

Dabei ist $\frac{\partial m}{\partial t}$ die Masse, welche pro Zeiteinheit t durch die Fläche A (z.B.) Membran, und $\frac{\partial c}{\partial x}$ der Konzentrationsgradient entgegen der Flußrichtung.Die Zahl D ist die Diffusionskonstante (abhängig von der Substanz).

Beispiel: Diffusion (zweites Fick'sches Gesetz) Aus dem ersten Fick'schen Gesetz und der Massenerhaltung ($\frac{\partial c}{\partial t} = -A\frac{\partial^2 m}{\partial x \partial t}$) folgt

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2}.$$

Beschreibende Statistik

Eine Messreihe

1	2	3	4	5	6	7	8	9	10	11	12	13	14
10.9	6.8	9.5	6.9	8.2	3.4	6.2	8.6	5.3	10.7	8.1	8.0	8.9	10.7

Wie kann man diese Daten geeignet darstellen?

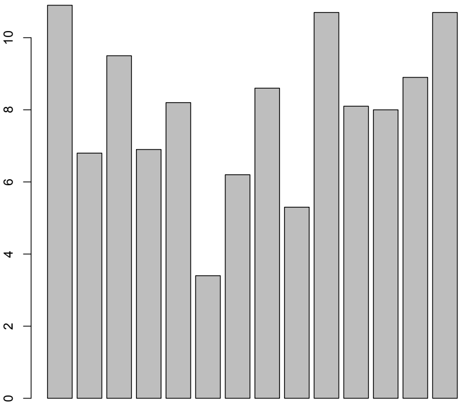
- Tabelle
- Plot (Barplot, Punktplot,...)
- Histogramm
- ...

Welche Informationen kann man aus diesen Daten ablesen?

- Häufigkeiten
- Mittelwert
- Streuung
- Ausreißer
- ...

Tabelle und Plot

1	2	3	4	5	6	7	8	9	10	11	12	13	14
10.9	6.8	9.5	6.9	8.2	3.4	6.2	8.6	5.3	10.7	8.1	8.0	8.9	10.7



Häufigkeiten

(Def.) Die absolute Häufigkeit $H(x)$ von x gibt an, wie oft x als Messwert auftritt. Die relative Häufigkeit von x ist $h(x) = H(x)/n$, wenn insgesamt n Messwerte vorliegen.

Im Beispiel:

1	2	3	4	5	6	7	8	9	10	11	12	13	14
10.9	6.8	9.5	6.9	8.2	3.4	6.2	8.6	5.3	10.7	8.1	8.0	8.9	10.7

$$H(10.9) = 1, \quad H(10.7) = 2, \quad H(4.0) = 0.$$

$$h(10.9) = \frac{1}{14}, \quad h(10.7) = \frac{1}{7}, \quad h(4.0) = 0.$$

Klassen, Häufigkeiten und Histogramme

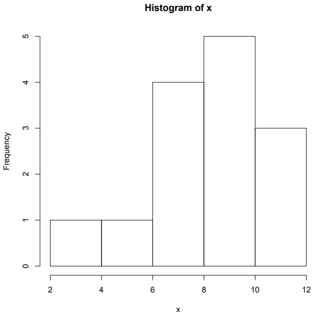
Absolute und relative Häufigkeiten können leicht in Tabellen dargestellt werden. Beispiel:

3.4	5.3	6.2	6.8	6.9	8.0	8.1	8.2	8.6	8.9	9.5	10.7	10.9
1	1	1	1	1	1	1	1	1	1	1	2	1

Zusammenfassen von Werten in Klassen:

Klasse	[2,4)	[4,6)	[6,8)	[8,10)	[10,12)
Häufigkeit	1	1	3	5	3

Darstellung von Häufigkeiten (!) als Histogramm:

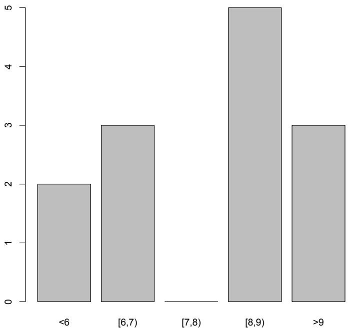


Klassen, Häufigkeiten und Histogramme

Andere Klasseneinteilung:

Klasse	< 6	[6,7)	[7,8)	[8,9)	> 9
Häufigkeit	2	3	0	5	3

Zugehöriges Histogramm:



- Wahl der Klasseneinteilung hängt davon ab, was man darstellen möchte
- Achsenbeschriftung!

Kenngrößen von Daten

(Def.) Seien $x_1, \ldots, x_n$ Messwerten/Daten. Das empirische Mittel von ist definiert als

$$\bar{\mu}_x := \frac{1}{n} \sum_{i=1}^{n} x_i$$

Der Median von $(x_1, \ldots, x_n)$ ist definiert als der Wert in der Mitte der aufsteigend geordneten Liste der Messwerte.

Das p-te Quantil ist dasjenige x_i, für welches ein Anteil p der Messwerte in der geordneten Liste links von x_i liegen, also kleiner sind. Perzentile sind Quantile mit $p = 0.1, 0.2, 0.3 \ldots$

Kenngrößen: Beispiel

1	2	3	4	5	6	7	8	9	10	11	12	13	14
10.9	6.8	9.5	6.9	8.2	3.4	6.2	8.6	5.3	10.7	8.1	8.0	8.9	10.7

- Empirisches Mittel: 8.014

Geordnete Daten:

3.4	5.3	6.2	6.8	6.9	8.0	8.1	8.2	8.6	8.9	9.5	10.7	10.7	10.9

2 Werte in der Mitte (da $n = 14$ gerade): 8.1, 8.2.
Median in diesem Fall $\frac{8.1+8.2}{2} = 8.15$.

Perzentile: Beispiel

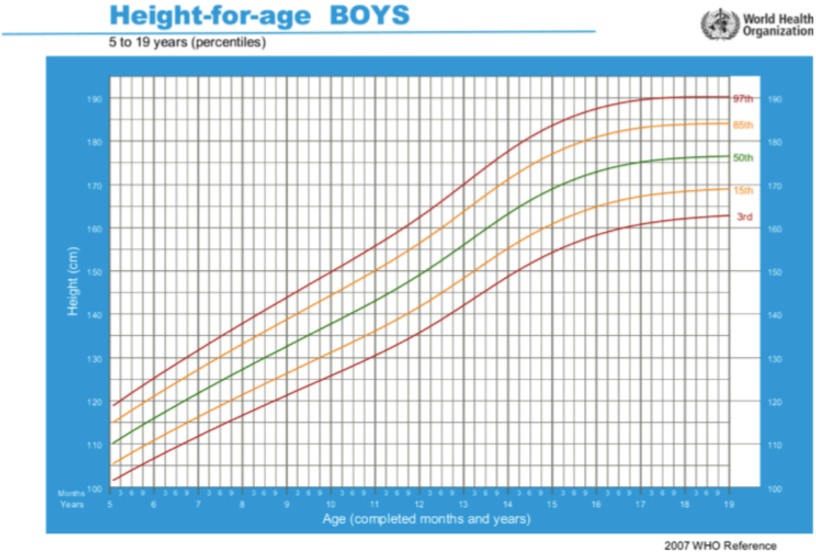


Abbildung 19: Wachstum von Jungen im Alter von 5 bis 19 Jahren

Kenngrößen von Daten

(Def.) Seien $x_1, ..., x_n$ Messwerten/Daten. Die empirische Varianz ist definiert als

$$\bar{s}_x^2 := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{\mu}_x)^2$$

Die empirische Standardabweichung $\bar{s}$ ist die Wurzel aus der empirischen Varianz.

- Empirisches Mittel: Durchschnittswert
- Empirische Varianz: Maß für die Streuung

Excel-Befehle:

- =Mittelwert(Wertebereich)
- =Median(Wertebereich)
- =Varianz(Werteberich), =Stabw(Wertebereich)
- =Quantil(Wertebereich, p) (z.B. $p = 0.1$)

Relative Häufigkeiten und Wahrscheinlichkeiten

Intuition: Wenn die Zahl der Messungen n groß ist, so ist die relative Häufigkeit eines Messwertes $h_n(x)$ ungefähr gleich der theoretischen Wahrscheinlichkeit, dass dieser Messwert auftritt. Formal:

$$\lim_{n \to \infty} h_n(x) = \mathbb{P}(x),$$

bzw.

$$h_n(x) \approx \mathbb{P}(x).$$

Das $\mathbb{P}$ steht für Wahrscheinlichkeit (probability, probabilitas).

Dies nennt man auch (empirisches) Gesetz der großen Zahlen.

Relative Häufigkeit (aus Messungen) als Schätzer für die Wahrscheinlichkeit.

Ereignisse und Mengen

Ein einzelner Messwert ist eine Zahl $x \in \mathbb{R}$, welche als Ergebnis eines zufälligen Vorgangs interpretiert wird.

Beispiel: Von n zufällig ausgewählten Probanden werden Alter x, Gewicht y und (systolischer) Blutdruck z gemessen.

Von Interesse sind Ereignisse, welche aus diesen Daten gebildet werden können. Dies sind Mengen:

- Der Proband ist mindestens 60: $A = \{x : x \geq 60\}$
- Der Proband wiegt unter 80 kg: $B = \{y : y < 80\}$
- Der systolische Blutdruck liegt zwischen 125 und 140: $C = \{z : z \in [125, 140]\} = \{z : 125 \leq z \leq 140\}$.

Davon abgeleitete Ereignisse:

- $A \cap B$: Ereignis A und B treten ein, also der Proband ist mindestens 60 Jahre alt und wiegt höchstens 80 kg.
- $A \cup B$: Ereignis A oder B (oder beide) treten ein
- A^c das Gegenteil von A tritt ein
- $(A \cup B) \cap C^c \ldots$