

Geodatenanalyse I:

Schließende Statistik und Wahrscheinlichkeiten

Kathrin Menberg



Stundenplan

	08:30 – 12:30 Uhr	13:30 – 17:30 Uhr
Montag	Tag 1 / Block 1	Tag 1 / Block 2
Dienstag	Tag 2 / Block 1	Tag 2 / Block 2
Mittwoch	Tag 3 / Block 1	Tag 3 / Block 2
Donnerstag	Tag 4 / Block 1	Tag 4 / Block 2
Freitag	Tag 5 / Block 1	Tag 5 / Block 2

- ▶ 2.1 Einführung und Deskriptive Statistik
- ▶ 2.2 Statistischen Testen
- ▶ **2.3 Schließende Statistik und Wahrscheinlichkeiten**

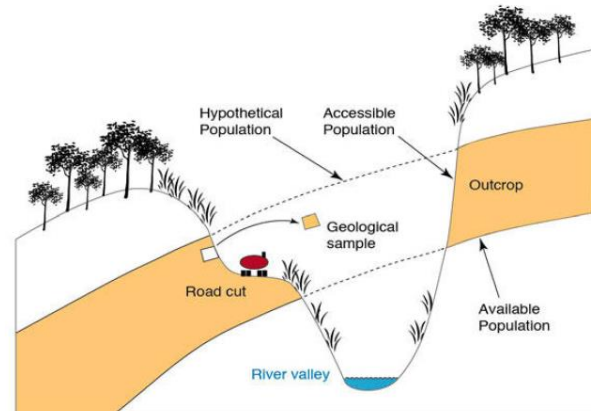
Lernziele Block 2.3

Am Ende der Stunde werden die Teilnehmer:

- ▶ ... verschiedene theoretische Verteilungen und deren statistische Momente kennen.
- ▶ ... Verteilungen an Datensätze anpassen und die Übereinstimmung bewerten und diskutieren können.
- ▶ ... mit den Grundlagen der Wahrscheinlichkeitsrechnung vertraut sein.

Anknüpfung

- ▶ Übung 1: Charakterisierung von Stichproben anhand von statistischen Parametern
- ▶ ... nun schauen wir uns die Verteilung der Grundgesamtheit an
- ▶ Annahme: $n \rightarrow \infty$
- ▶ Schließende Statistik
- ▶ Wahrscheinlichkeit



Trauth (2015) (Fig. 1.1)

Was ist Wahrscheinlichkeit?

- ▶ Relative Häufigkeit in Zufallsexperimenten
- ▶ Zufallsexperiment = Vorgang, der beliebig oft unter den gleichen Bedingungen wiederholbar ist
- ▶ ... und dessen Ausgang nicht mit Sicherheit vorhergesagt werden kann

Beispiel:

- ▶ Medikament, dass bei 80% der Patienten wirkt
- ▶ Wahrscheinlichkeit der Wirkung bei zufällig herausgegriffenem Patienten $p = 0.8$

Rechnen mit Wahrscheinlichkeiten

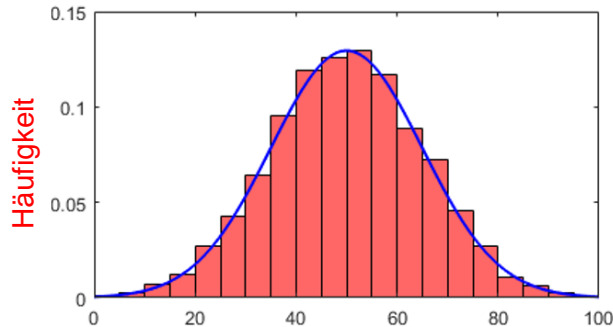
- ▶ Drei Grundregeln (nach A. Kolmogorov):
 - ▶ Wahrscheinlichkeit als reelle, nichtnegative Zahl: $p(A) \geq 0$
 - ▶ Sicheres Ereignis hat Wahrscheinlichkeit 1: $p(S) = 1$
 - ▶ Wenn sich A und B ausschließen gilt: $p(A + B) = p(A) + p(B)$
- ▶ Bedingte Wahrscheinlichkeit (conditional probability)
 - ▶ Wahrscheinlichkeit für Ereignis A , unter der Bedingung ein Ereignis B sei eingetreten: $p(A|B)$
- ▶ Totale Wahrscheinlichkeit
 - ▶ Wahrscheinlichkeit für Ereignis A ergibt sich aus den bedingten Wahrscheinlichkeiten und den Wahrscheinlichkeiten dafür dass die Bedingungen eintreten: $p(A) = \sum_i p(A|B_i) p(B_i)$

Wahrscheinlichkeitsverteilungen

- ▶ Deskriptive Statistik: Stichprobe → Messwert
- ▶ Schließende Statistik: Zufallsgröße → Wahrscheinlichkeit

Deskriptive Statistik

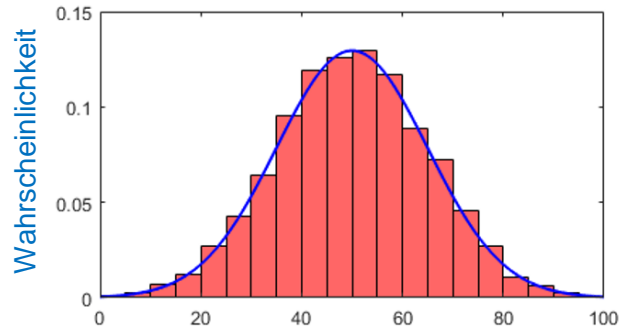
Häufigkeitsverteilung



Gesamtheit der Verteilung ergibt
Anzahl der Stichproben

Schließende Statistik

Wahrscheinlichkeitsverteilung

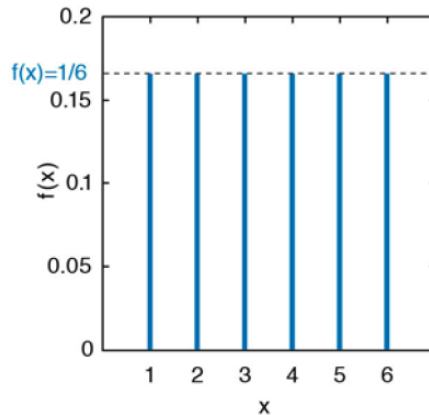


Gesamtheit der Verteilung ergibt
Wahrscheinlichkeit $p = 1$

Theoretische Verteilungen

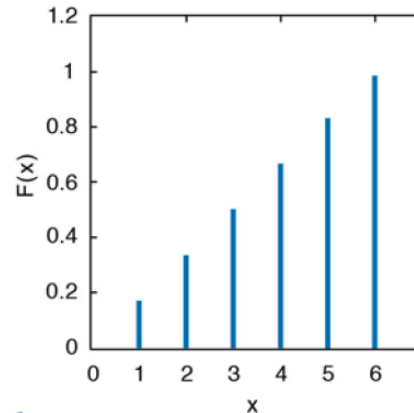
- Diskrete Werte: Wahrscheinlichkeitsfunktionen (probability mass function)
- Uniformverteilung, Gleichverteilung (Minimum, Maximum)

Wahrscheinlichkeitsfunktion $f(x)$



a

Kumulative
Wahrscheinlichkeitsfunktion $F(x)$

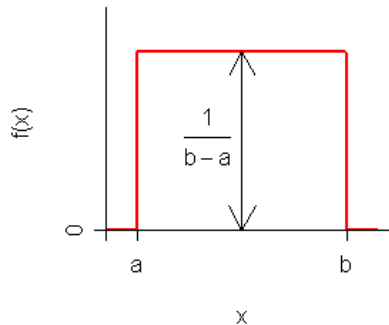


b

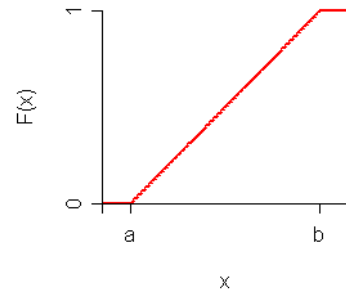
Theoretische Verteilungen

- ▶ Stetige Werte: Wahrscheinlichkeitsdichtefunktionen (probability density function)
- ▶ Uniformverteilung, Gleichverteilung (Minimum, Maximum)
 - ▶ $U(\min, \max)$

Dichtefunktion $f(x)$

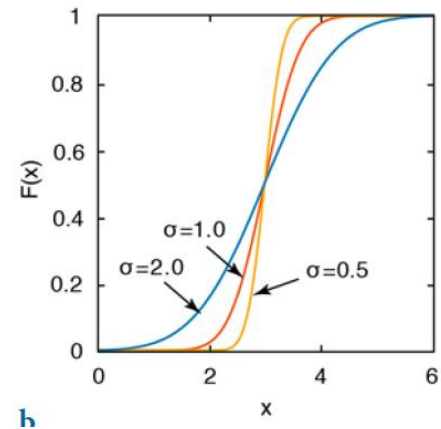
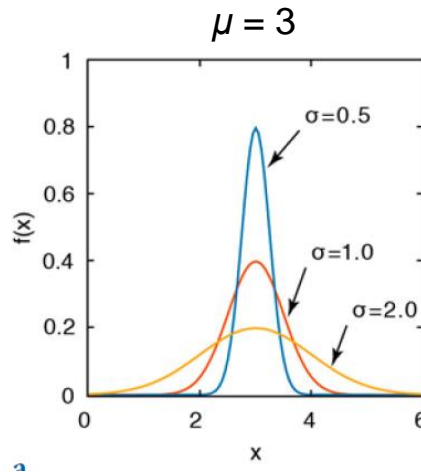


Verteilungsfunktion $F(x)$



Häufig verwendete Verteilungen

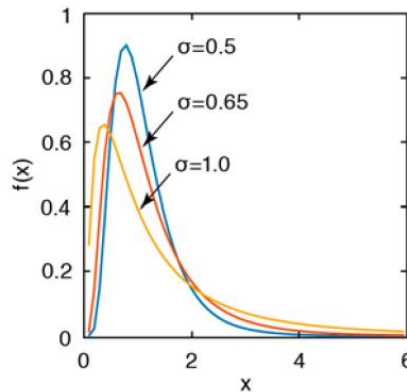
- ▶ Normal, -Gaußverteilung (Mittelwert, Varianz)
 - ▶ $N(\mu, \sigma^2)$
 - ▶ Mean = Median = Mode
 - ▶ Skewness = 0
 - ▶ Kurtosis = 3
 - ▶ $x \in (-\infty, +\infty)$



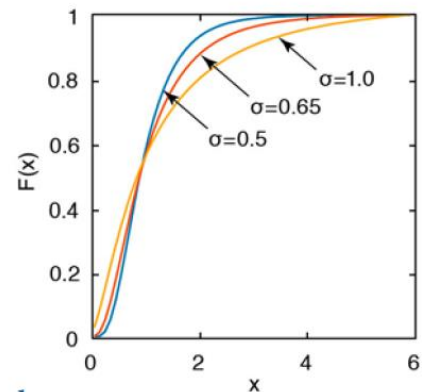
Trauth (2015) Fig. 3.7

Häufig verwendete Verteilungen

- ▶ Log-Normalverteilung (Mittelwert $\mu_{\log n}$, Standardabweichung $\sigma_{\log n}$)
 - ▶ Mean \neq Median \neq Mode
 - ▶ Skewness > 0
 - ▶ $x > 0$



a



b

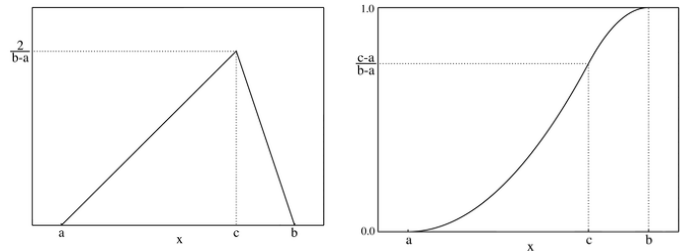
Trauth (2015) Fig. 3.7

Häufig verwendete Verteilungen

► Triangularverteilung (min, mode, max)

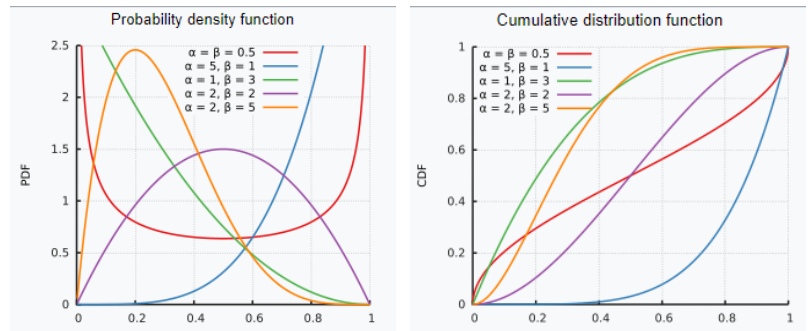
► Mean \neq Median \neq Mode

► $x \in (\min, \max)$



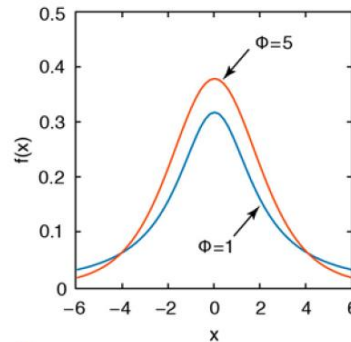
► Betaverteilung (α, β)

► $x \in (0, 1)$

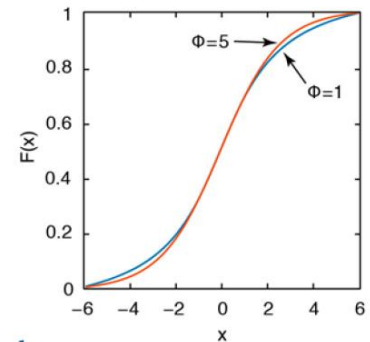


Spezielle theoretische Verteilungen

- Studentsche t-Verteilung
(Freiheitsgrade, Φ)

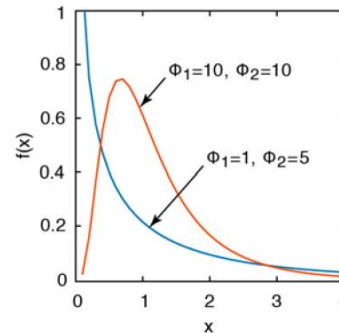


a

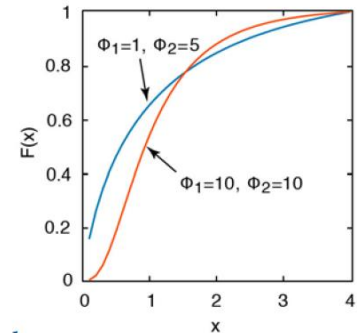


b

- Fisher Verteilung
(Freiheitsgrade, Φ_1 , Φ_2)



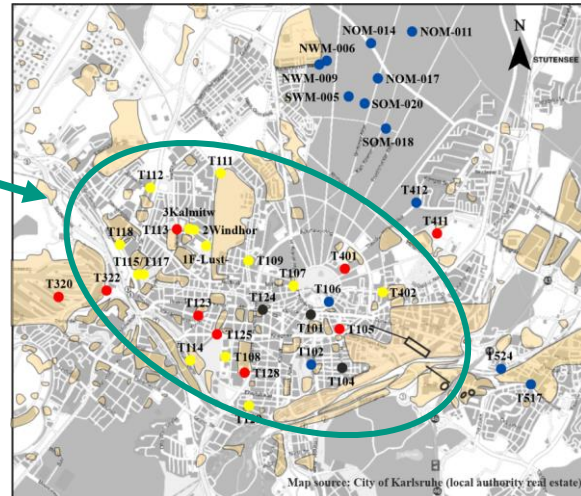
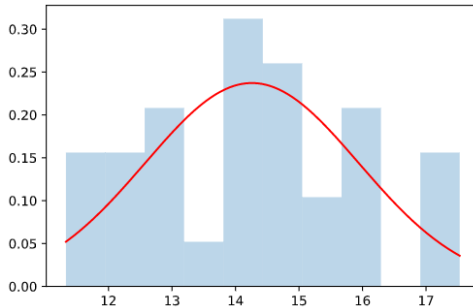
a



b

Übung 2.3: Schließende Statistik

- Grundwasserdatensatz Karlsruhe
- Anpassung theoretische Verteilung an gemessene Stichproben



- Aufgaben in Jupyter Notebook: geodatenanalyse_1-2-3

Aufgabenbesprechung

- ▶ Anpassung Normalverteilung an Grundwassertemperaturen
 - ▶ $\text{mean_fit} = 14.25^{\circ}\text{C}$, $\text{variance_fit} = 1.6$
- ▶ Zufallswerte mit $n = 50$
 - ▶ $\text{mean_sample} = 14.43$, $\text{variance_sample} = 2.8$
- ▶ Zufallswerte mit $n = 500,000$
 - ▶ $\text{mean_sample2} = 14.25$, $\text{variance_sample2} = 2.8$
 - ▶ $\text{min} = 6.42^{\circ}\text{C}$, $\text{max} = 22.37$
- ▶ Gestutzte Normalverteilung
 - ▶ $\text{lower_bound} = 12$, $\text{upper_bound} = 18$
 - ▶ $\text{Min} = 11.33^{\circ}\text{C}$, $\text{max} = 19.11^{\circ}\text{C}$

Literatur

- ▶ Trauth (2015) MATLAB Recipes for Earth Sciences (4th Ed.), Springer
- ▶ Tschirk (2014) Statistik: Klassisch oder Bayes, Springer
- ▶ Koch et al. (2020) Groundwater fauna in an urban area: natural or affected?, Hydrology and Earth System Sciences Discussions

Nützliche Weblinks:

- ▶ Copluas: <https://twiecki.io/blog/2018/05/03/copulas/>

