

Geodatenanalyse I: Einführung und Deskriptive Statistik

Kathrin Menberg



Stundenplan

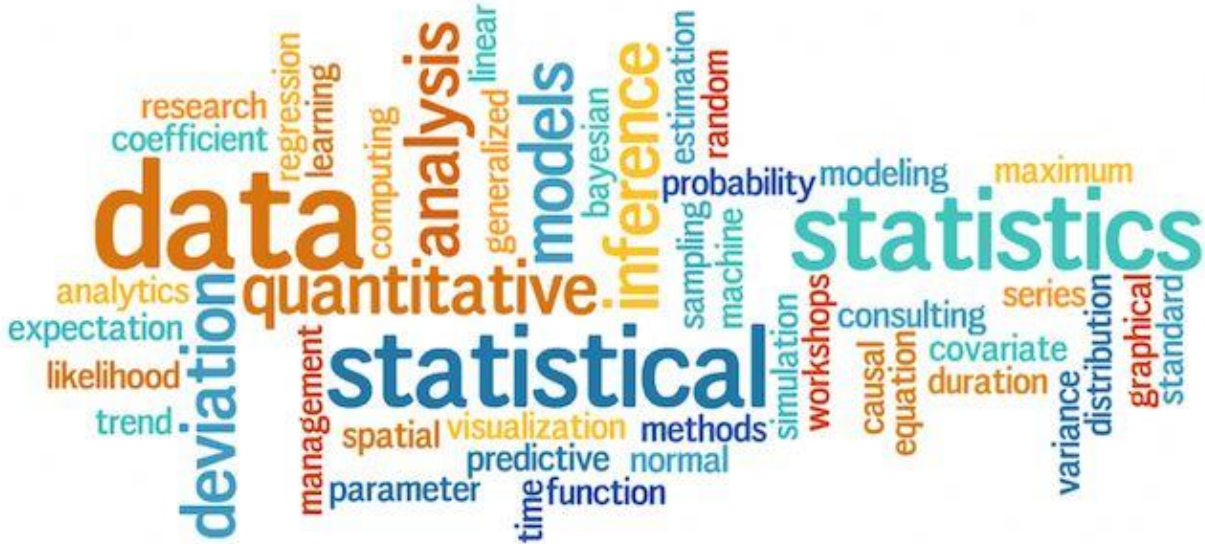
	08:30 – 12:30 Uhr	13:30 – 17:30 Uhr
Montag	Tag 1 / Block 1	Tag 1 / Block 2
Dienstag	Tag 2 / Block 1	Tag 2 / Block 2
Mittwoch	Tag 3 / Block 1	Tag 3 / Block 2
Donnerstag	Tag 4 / Block 1	Tag 4 / Block 2
Freitag	Tag 5 / Block 1	Tag 5 / Block 2

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

Lernziele Block 2.1

Am Ende der Stunde werden die Teilnehmer:

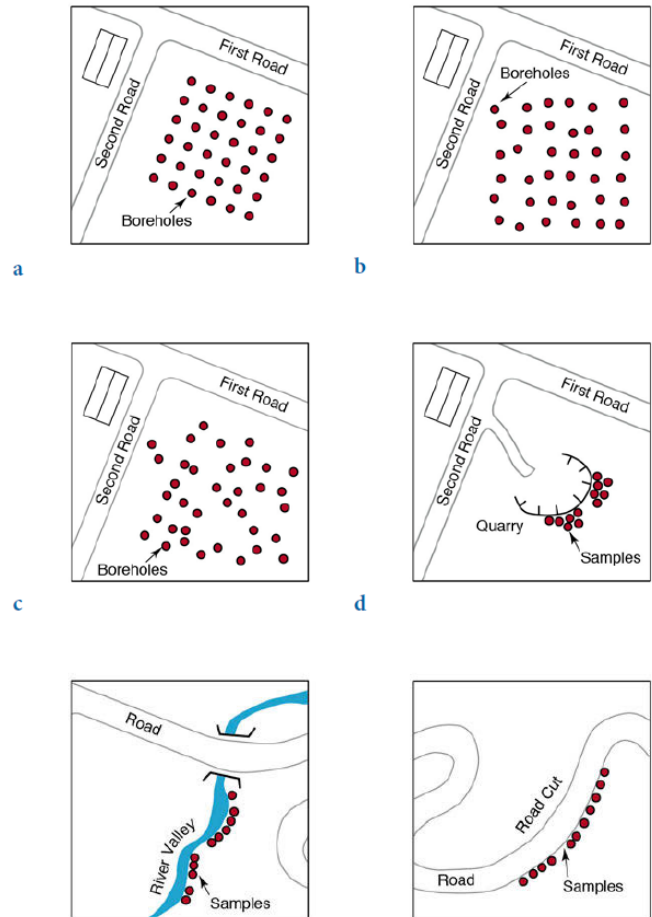
- ▶ ... grundlegende Begriffe der univariaten Statistik und Datenanalyse kennen.
- ▶ ... in Python statistische Momente bestimmen können.
- ▶ ... empirische Verteilungen charakterisieren können.



Geostatistik – auf Mathematik basierende Methoden zur Analyse quantitativer Daten mit Raumbezug (Geodaten)

Geodaten

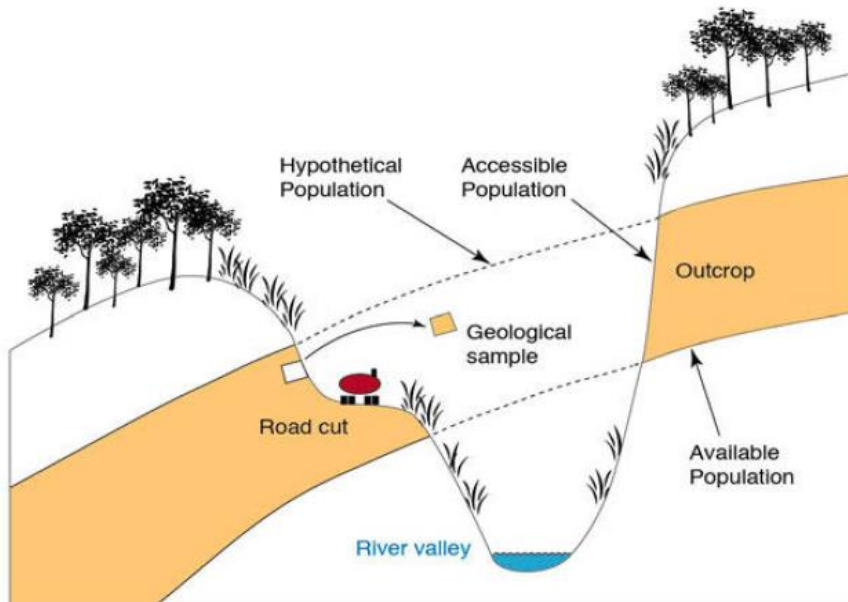
- ▶ Geschätzt 80% aller Daten haben einen Raumbezug!
- ▶ Sammeln von Daten (Feld, Labor, Satellitendaten, usw.)
 - ▶ Begrenzte Probenzahl (n)
 - ▶ Messunsicherheit



Trauth (2015) Fig. 1.2

► Grundgesamtheit und Stichprobe

(engl. population, sample)



Trauth (2015) (Fig. 1.1)

Typische Arten von Geodaten

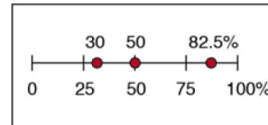
Nominale Daten (nominal data)

Cyclotella ocellata
C. meneghiniana
C. ambigua
C. agassizensis
Aulacoseira granulata
A. granulata var. curvata
A. italica
Epithemia zebra
E. sorex
Thalassioseira faurii

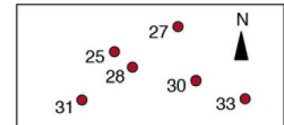
Ordinale Daten (ordinal data)

1. Talc
2. Gypsum
3. Calcite
4. Flurite
5. Apatite
6. Orthoclase
7. Quarz
8. Topaz
9. Corundum
10. Diamond

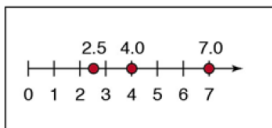
Geschlossene Daten (closed data)



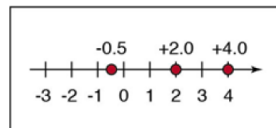
Räumliche Daten (spatial data)



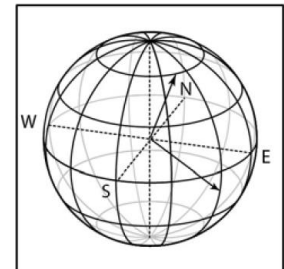
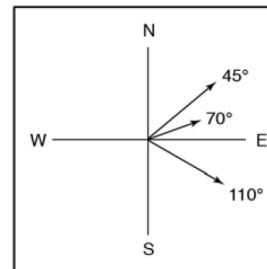
Verhältnisdaten (ratio data)



Intervalldaten (interval data)



Richtungsabhängige Daten (directional data)



Trauth (2015) Fig. 1.3

Weitere Grundbegriffe

- ▶ Messgröße (measured variable)
- ▶ Zufallsgröße (random variable)
- ▶ Diskrete Daten, bzw. Funktionen (discrete data)
- ▶ Stetige Daten, bzw. Funktionen (continuous data)
- ▶ Parameter
- ▶ Variable
- ▶ Freiheitsgrade (degrees of freedom)
- ▶ Wahrscheinlichkeit (probability)
- ▶ Unsicherheit (uncertainty)

Diskrete Daten



Stetige Daten

Methoden zur Datenanalyse

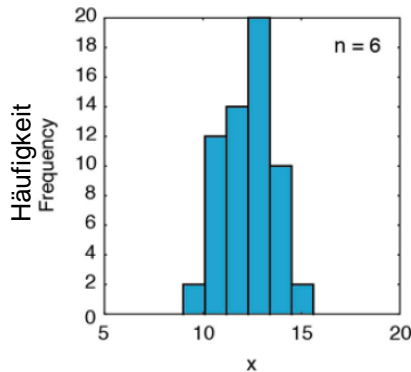
- ▶ Univariate Methoden: eine unabhängige Messgröße
- ▶ Bivariate Methoden: zwei abhängige Messgrößen
- ▶ Multivariate Methoden: mehrdimensionale Datensätze
- ▶ Zeitreihenanalyse: Datenwerte als Funktion der Zeit
- ▶ Räumliche Analyse: Daten mit Koordinaten in 2D oder 3D

- ▶ ... und viele mehr.

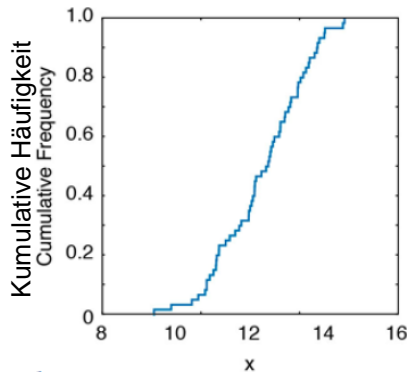
Beschreibende Statistik

- ▶ Statistische Charakterisierung von Stichproben
- ▶ Empirische Verteilungen von Messwerten (empirical distribution)
- ▶ Graphische Darstellung

Empirische Häufigkeitsverteilung



Kumulative Häufigkeitsverteilung



Trauth (2015) (Fig. 3.1)

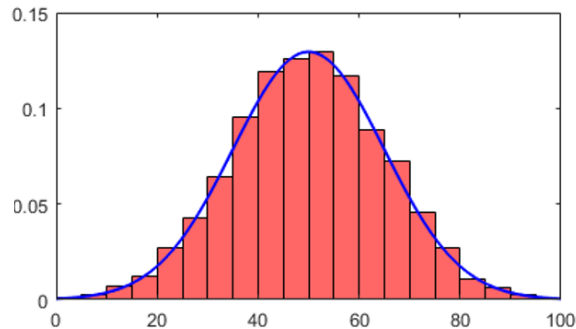
- ▶ ... auch deskriptive Statistik genannt (descriptive statistics)

Schließende Statistik

- ▶ Analyse der Grundgesamtheit
- ▶ Theoretische Verteilungen von Zufallsvariablen (theoretical distribution)

Empirische Häufigkeitsverteilung

Theoretische (angepasste) Wahrscheinlichkeitsverteilung



- ▶ ... dazu später mehr!

Charakterisierung von Stichproben

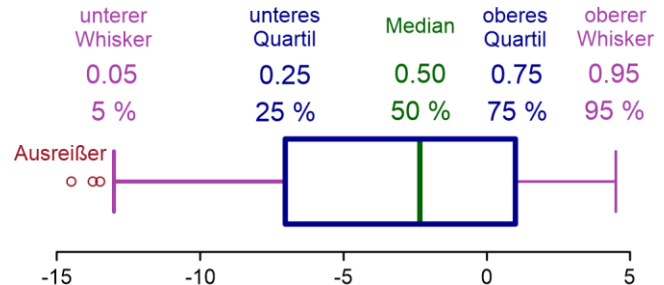
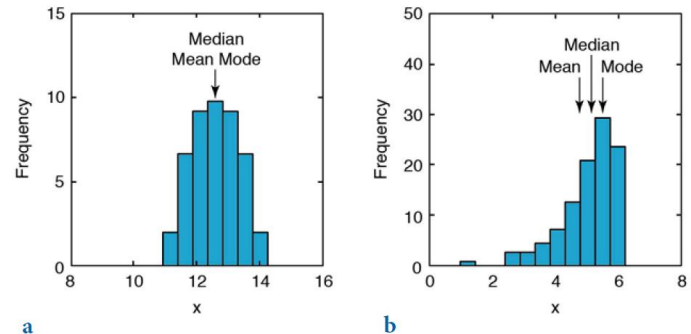
Statistische Parameter (statistical measures)

Trauth (2015) (Fig. 3.2)

1. Lageparameter (central tendency)

- ▶ Arithmetisches Mittel (mean)
- ▶ Geometrisches Mittel
- ▶ Harmonisches Mittel

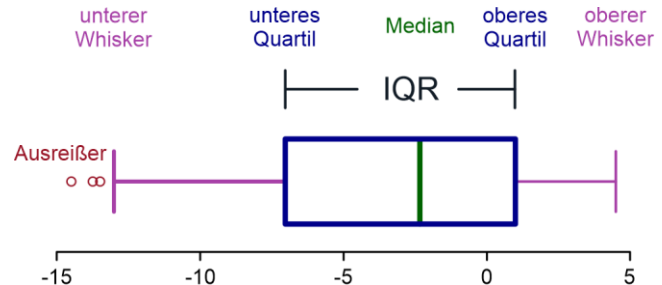
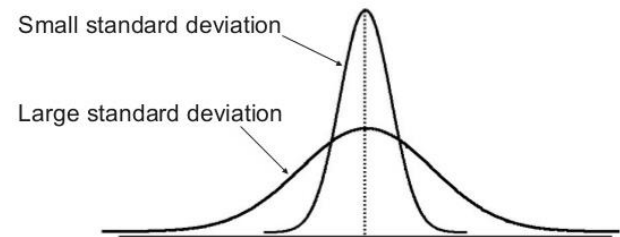
- ▶ Median (median)
- ▶ Modus (mode)
 - ▶ Nur für diskrete, bzw. nominale Daten!
- ▶ Quartile, Quantile, Perzentile
- ▶ usw.



Charakterisierung von Stichproben

2. Streuungsmaß (dispersion)

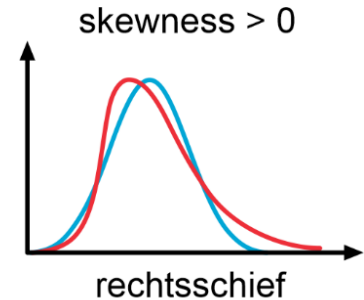
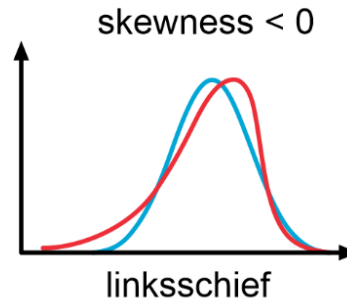
- ▶ Range (Maximum – Minimum)
- ▶ empirische Varianz (σ^2)
- ▶ empirische Standardabweichung (σ)
- ▶ (Inter)Quartilabstand (IQR)
- ▶ usw.



Charakterisierung von Stichproben

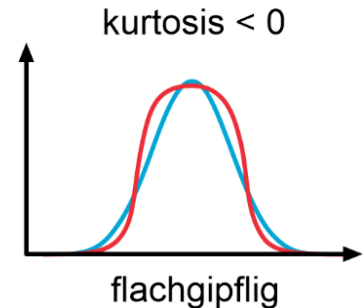
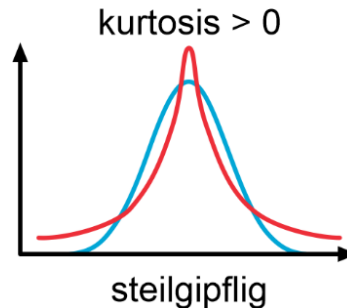
3. Schiefe (skewness)

- ▶ nach Pearson
- ▶ nach Fisher
- ▶ Quantil-basiert
- ▶ usw.



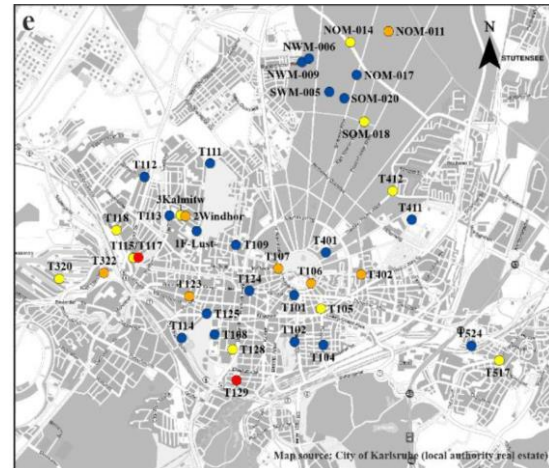
4. Wölbung (kurtosis)

- ▶ Im Vergleich zu einer Normalverteilung
- ▶ nach Fisher



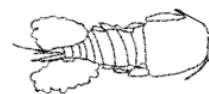
Übung 2.1: Deskriptive Statistik

- ▶ Grundwasserdatensatz Karlsruhe
- ▶ Messwerte zu
Grundwassertemperatur,
hydrochemische Parameter,
faunistische Daten
- ▶ Datentypen
- ▶ Bestimmung statistischer Parameter
- ▶ Aufgaben in Jupyter Notebook:
geodatenanalyse_1-2-1



Koch et al. (2020) HESS-D

Crustaceen



Oligochaeten



Aufgabenbesprechung

Parameter	Datentyp
Pegel	Nominale Daten
Tiefe	Verhältnisdaten (stetig)
Sauerstoff	Verhältnisdaten (stetig)
Temperatur	Intervalldaten (stetig)
Elektrische Leitfähigkeit	Verhältnisdaten (stetig)
pH Wert	Verhältnisdaten (stetig)
Eisen, Mangan, Phosphat, Nitrat	Verhältnisdaten (stetig)
Detritus, Sediment	Ordinale Daten (diskret)
Geologische Einheit, Flächennutzung	Ordinale Daten (diskret)
Anzahl Arten, Anzahl Individuen	Verhältnisdaten (stetig)
Anteil Crustaceen, Anteil Oligochaeten	Geschlossene Daten (stetig)

Aufgabenbesprechung

Variable	Python-Datentyp
GWT	list
n	int

Variable	Wert
arithm. Mittel	13.5
mean	13.5
Median_1	14.1
Median_2	14.0
Mode (Geologie) (nur für diskrete Daten!)	4
Quartile	[11.4, 14.0, 15.0]
Range	7.0
IQR	3.6
Standardabweichung	2.11
Varianz	4.45

Aufgabenbesprechung

Variable	Python-Datentyp
Skewness Pearson	-0.27
Skewness Fisher	0.29
Kurtosis	3.53

Literatur

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

