

Quantitatives wissenschaftliches Arbeiten



<https://iuk.one/1012-1050>

Clemens H. Cap

ORCID: 0000-0003-3958-6136

Department of Computer Science
University of **Rostock**
Rostock, Germany
clemens.cap@uni-rostock.de



3

Wissenschaftliches Arbeiten in Ingenieur- und Naturwissenschaften ist sehr oft ein **quantitatives** Arbeiten.

Quantisierung und Operationalisierung sichern **Nachvollziehbarkeit** und **Präzision**.

Wir geben hier einen (kurzen) Abriß über einige Aspekte quantitativen Arbeitens.

3 Arten von Skalen

Nominale Skala: Unterscheidbare Ausprägungen ohne weitere Struktur.
Bsp: Blutgruppe (A, B, AB, 0), Protokoll (TCP, UDP).

Ordinale Skala: Die Ausprägungen erlauben eine Reihung (Ordnung).

- **Transitiv:** Wenn $a > b$ und $b > c$ dann auch $a > c$.
- **Azyklisch:** Es gibt keine Zyklen: $a_1 > a_2, a_2 > a_3 \dots a_{n-1} > a_n$ und $a_n > a_1$.
- **Vergleichbar:** Für zwei verschiedene Objekte a und b gilt $a > b$ oder $b > a$.

Ordnungs-Strukturen sind **stets** transitiv und azyklisch.

Hinsichtlich der Vergleichbarkeit bestehen verschiedene Begrifflichkeiten.

Bsp: Leistungsumfang eines Textverarbeitungssystems (\LaTeX > Word > emacs)

Metrische Skala: Ausprägungen erlauben die Definition einer Maßzahl.

Bsp: Elektrischer Widerstand (3 [Ω]), Speichergröße (10 [GBit]), Uhrzeit (3:24:12).

Intervall Skala: Ausprägungen erlauben Abstandsbegriff aber keinen Nullpunkt.

Bsp: Falsch: Der Algorithmus terminierte um 3:15 Uhr.

Bsp: Richtig: Der Algorithmus terminierte nach 2 Minuten.

Verhältnis Skala: Ausprägungen erlauben einen Nullpunkt.

Rechenoperationen, Durchschnitte und quantitative Vergleiche sind sinnvoll.

Bsp: Richtig: Die Datenrate beträgt 20 [Mbit/s].

Bsp: Richtig: Die durchschnittliche Rechenzeit betrug 2 [min].

Bsp: Falsch: Die durchschnittliche Postleitzahl eines Studenten in Mecklenburg?

Leitsatz

Rechenoperationen und Vergleiche dürfen nur bei einem **tieferen Verständnis** der Operationalisierung und des inneren Zusammenhangs der Skala gebildet werden.

Beispiel: Sendeleistung bei drahtlosen Kommunikationsverfahren

Leistung kann angegeben sein:

- In der linearen Einheit [mW] Milli-Watt.
- In der logarithmischen Einheit [dBm] Dezibel-Milliwatt.

Konkrete Zahlenwerte:

- Alice sendet mit 10 [mW] = 10 [dBm].
- Bob sendet mit 1000 [mW] = 30 [dBm].

Fragen:

- Sendet Alice mit der 100-fachen oder mit der 3-fachen Leistung?
- Macht ein arithmetisches Mittel in der [mW] Skala Sinn?
- Macht ein arithmetisches Mittel in der [dBm] Skala Sinn?

Kontinuierliche und Diskrete Skalen (1)

Diskrete Skala: Endlich $\{a_1, a_2, \dots, a_n\}$ oder abzählbar $\{a_1, a_2, \dots\}$ viele Werte.

Kontinuierliche Skala: Stetig-überabzählbar-unendlich viele Werte. Bsp: \mathbb{R} .
Wertangaben empirischer Wissenschaften typisch fehlerbehaftet. (Gegenbeispiel: π).

Werkzeuge zur Notation der Präzision auf kontinuierlichen Skalen:

- **Klammerung:** Eingeklammerte Werte sind unsicher: 3.045(2)
- **Plus-Minus:** Angabe möglicher Schwankungsbreite: 3.045 ± 0.0003
- **Intervall:** Angabe des sicheren Intervalls [0.34, 0.39]
- **Konfidenzintervall:** Angabe des wahrscheinlichen Intervalls: Zu 95% in [0.3, 0.5]
- **Verteilung:** Durch Form & Parameter: Normal-verteilt mit $\mu = 3, \sigma = 0.46$
- **Verteilung:** Durch kumulative Verteilungsfunktion oder Dichte-Funktion.

Leitsatz: Adäquatheit der Präzision und der Darstellung

Bei kontinuierlichen Wertebereichen muß auf die **Adäquatheit** der benutzten Angaben und Darstellungen geachtet werden.

Beispiele für kontinuierliche Skalen (1)

Situation 1: Digitale Daten vom Rechner übernehmen.

Wegen Computerarithmetik werden *immer* 64 (Wortbreite) Bits ausgegeben.

Bsp: Timer sagt: 0.345 242 959 121 [ms]. Ist das wirklich Femto-Sekunden genau?

Daher: Muß echte Genauigkeit ermitteln, Angabe anpassen, Gründe dokumentieren.

Situation 2: Anzahl der Messungen anpassen.

Bei unpräzisen Wertebereichen ist eine **Einzelmessung nicht ausreichend**.

Daher: Führe mehrere Messungen durch und werte diese aus!

Gib an: **Minimal:** Anzahl der Messungen, **Mittelwert** und **Standardabweichung**.

Maximal: **Komplette Verteilung** aller Meßwerte.

Wenn (fast) immer (fast) derselbe Wert herauskommt,
dann beschreibt man das kurz und spart sich *danach* den Aufwand.

Beispiele für kontinuierliche Skalen (2)

Situation 3: Visualisieren von Datenpunkten.

Es wird in $n = 10$ Situationen eine Variablen y ermittelt.

Ist Darstellung als Kurve $y(x)$ oder $y(t)$ mit interpolierten Zwischenwerten adäquat?

Nein, wenn die echten 10 Meßpunkte nicht separat markiert sind.

Nein, wenn y sachlich nicht von kontinuierlicher Variablen abhängt.

Nein, wenn das nur n unabhängige, diskrete Meßpunkte darstellen soll.

Ja, wenn es einen **sinnvollen** Grund für die Interpolation gibt und dieser im begleitenden Text auch dargestellt wird.

Bsp: 10 Startupzeiten ein und desselben Programms.

Ja, wenn Messungen nacheinander geschehen und Caching-Effekt möglich.

Nein, wenn das "nur" wiederholte Messungen an verschiedenen Tagen sind.

Beispiel: Metrische Interpretation einer Nominal-Skala

Konkret: Es wurden $n = 10$ Messungen gemacht.

Diese werden nun als Zeitverlauf dargestellt, obwohl der Messung kein Zeitverlauf zugrunde liegt.

Beispiel: Metrische Interpretation einer Ordinalskala.

Konkret: Schulnoten und Punkte.

- **Deutsche Uni-Noten** ("sehr-gut" bis "nicht-genügend") sind ordinale Merkmale.
Sie erlauben keine sinnvolle Durchschnittsbildung.
- **Punktezahlen** sind metrische Merkmale.
Sie erlauben Durchschnittsbildung.
- **Schweizer Schulnoten** sind metrisch definiert.
Sie erlauben sinnvolle Durchschnittsbildung.
Aber: Es gibt gleich viele negative Noten wie bestandene Noten.

... wobei fehlerhafte oder zweifelhafte Werte weggelassen wurden. (Aus einer meteorologischen Arbeit: p. 2, https://meteo.boku.ac.at/klima/berichte/wald_interpolation.pdf)

Was sind "fehlerhafte oder zweifelhafte Werte"?

- Sie erscheinen dem Autor wenig plausibel, weil sie seiner Theorie nicht entsprechen.
- Sie sind physikalisch unsinnig (Bsp: -500 [°Celsius] gibt es nicht).
- Sie sind anschaulich zweifelhaft (Bsp: +400 [°Celsius] Lufttemperatur in Berlin).
- Meßgeräte war abgeschaltet, aber Computer hat einen Datenpunkt empfangen.
- In der Datei standen die Werte 3.4, 3.7 und 1999999999.0
- In der Datei standen die Werte 3.4, 3.7 und BatteryEmpty.
- In der Datei standen die Werte 120, 126, -3, 124. Ev. ein Bereichsüberlauf?

Leitsatz

Läßt man Daten weg, dann erklärt man **genau** das **wie** und das **warum**.

Checkliste: Wie gestaltet man eine Abbildung?

Generell beachtet man

- 1 **Numerierung:** Damit Abbildung referenziert werden kann.
- 2 **Bildunterschrift:** Was sehen wir in der Abbildung.
- 3 **Rohdaten:** Wie entstanden? Wie bereinigt? Wo erhältlich?
- 4 **Beschriftungen:** Was bedeutet welche Achse, welches Symbole?
- 5 **Ticks:** Linear, logarithmisch, keine?
- 6 **Einheiten:** In welcher Einheit wird dargestellt.
- 7 **Legende:** Weitere Informationen, falls erforderlich.
- 8 **Fehler:** Wenn erforderlich Fehlerbalken oder Box Plot.

Eigene Wissenschaft – hier nur einige Anmerkungen.

Plot: Marker oder durchgezogen?

- Für theoretische Modelle: Durchgezogene Linie.
- Für Einzel-Experimente: Ein Marker pro Experiment.
- Für Experimental-Serie: Fehlerbalken oder Box-Plot

Die schwarzen Kugeln sind **Meßpunkte**.

Die horizontal begrenzten, vertikalen Linien geben den **Fehlerbalken** an.

Die durchgezogene Kurve ist eine **Interpolation**.

Legende erklärt die Bedeutung der Kurve.

\LaTeX Code Beispiel: <https://tex.stackexchange.com/questions/397190/>.

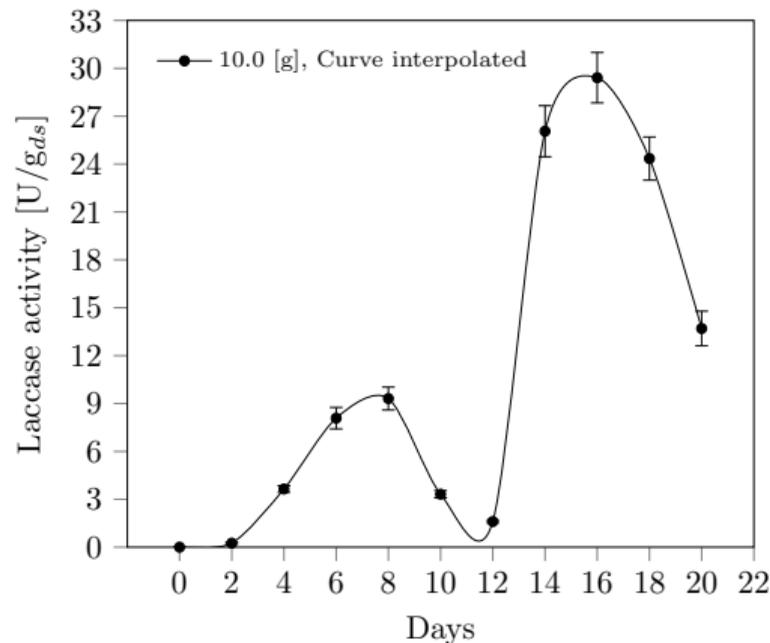


Abb. 1: Beispiel für einen Plot mit Error Bar

Interpolation (1)

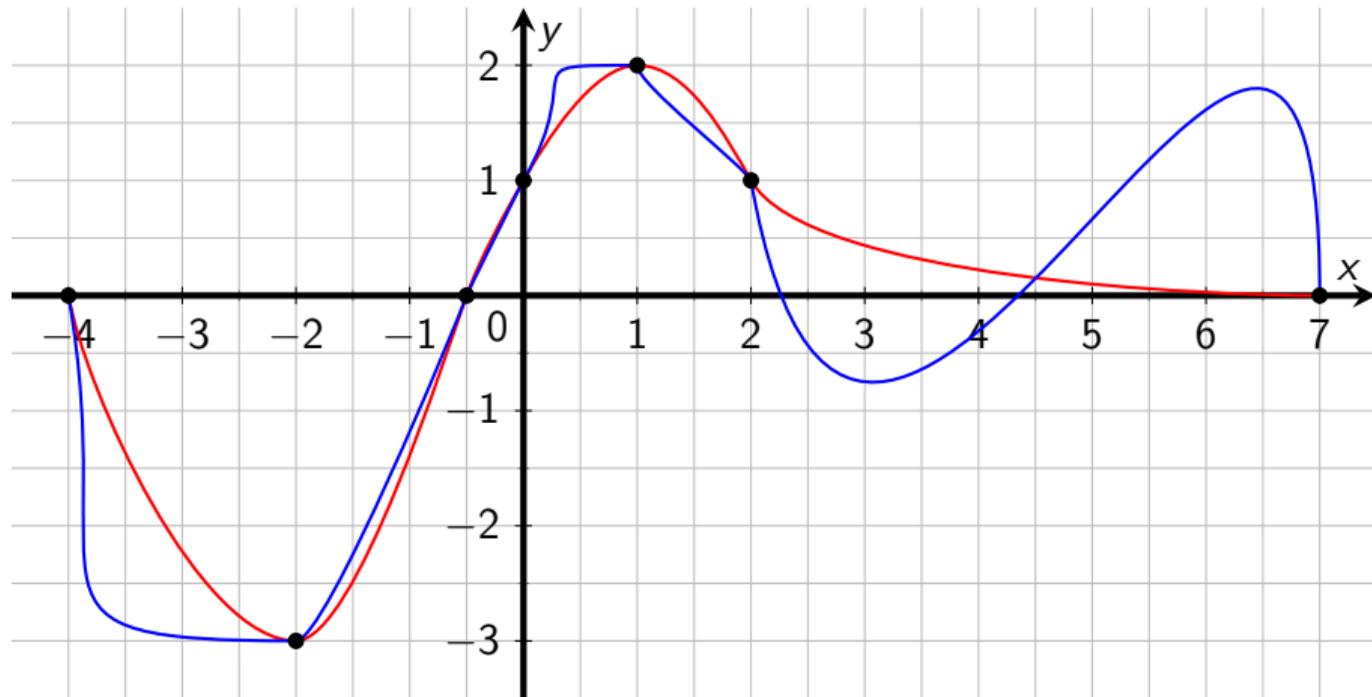


Abb. 2: Gemessen wurden die schwarzen Punkte. Interpolation kann daraus stark unterschiedliche Kurvenverläufe machen. Wenn es keine **sachliche Grundlage** für eine bestimmte Interpolation gibt, sollte man das besser bleiben lassen.

© Rechte siehe Anhang.

Interpolation

Interpolation ist die **Verallgemeinerung von Meßwerten** an einigen Stellen auf weitere Stellen, in denen keine Meßwerte vorliegen.

Sachgründe für eine Interpolation könnten sein:

- Vermute fast linearen Zusammenhangs – linear Interpolation sinnvoll.
- Vermute kleine Änderungsraten – polynomiale Interpolation sinnvoll.
- Vermute theoretisches Modell – Modellanpassung sinnvoll.

Leitsatz

Ohne sachliche Gründe ergibt Interpolation nur sinnfreie zufällige Werte und sollte daher besser unterbleiben.

Median: Teilt Meßreihe in zwei Hälften.

- Eine Hälfte ist größer als der Median.
- Eine Hälfte ist kleiner als der Median.

Quartil: Teilt Meßreihe in 1:3 bzw 3:1

Whisker: Oft: 2,5% und 97,5% Quartil.

Ausreißer: Außerhalb der Whisker.

Wichtig: **Gerne** anders, wenn Gründe.
Jedenfalls dokumentieren.

L^AT_EXCode Beispiel: <https://tex.stackexchange.com/questions/84541/>.

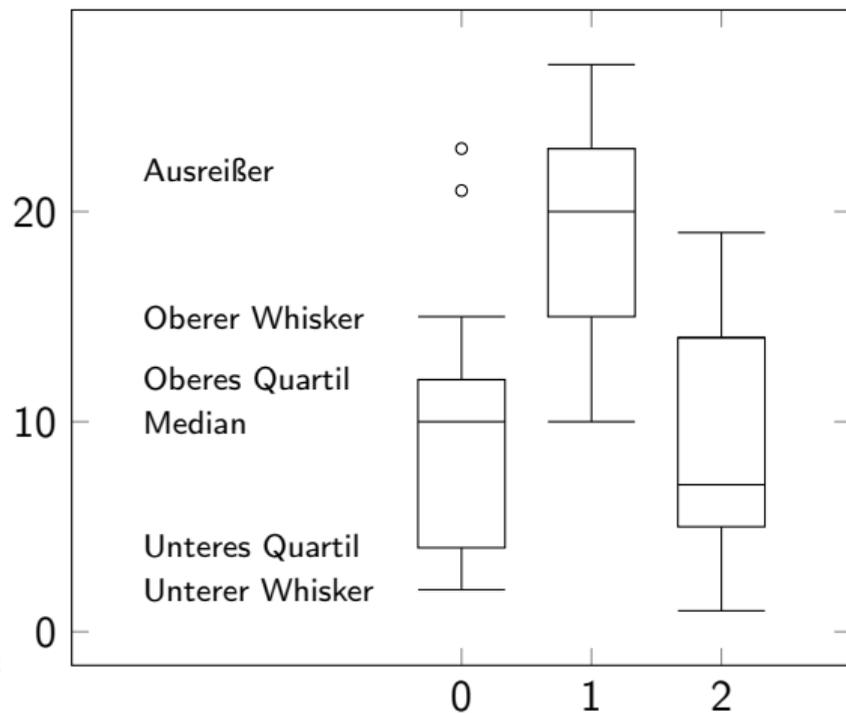


Abb. 3: Ein typischer Box-Plot.



Abb. 4: Rainbow Farbskala

Für dreidimensionale Plots und Konturen wird oft eine Farbskala benutzt.

Die beliebte **Regenbogen-Farbskala** soll **nicht** verwendet werden:

- Skalen-Abstufung bei gelb ändert sich viel rascher als im Grünen.
- Schlecht geeignet für farbenblinde Personen.
- Farbskala hat keine natürliche Semantik sondern ist erlernt.

Farbskalen (2)

Empfohlene Farbskalen:

- Siehe Abbildung 5.
- Siehe Farbmischer auf <https://colorbrewer2.org/>.
- [Kenneth Moreland: Why we use bad color maps and what you can do about it.](#)

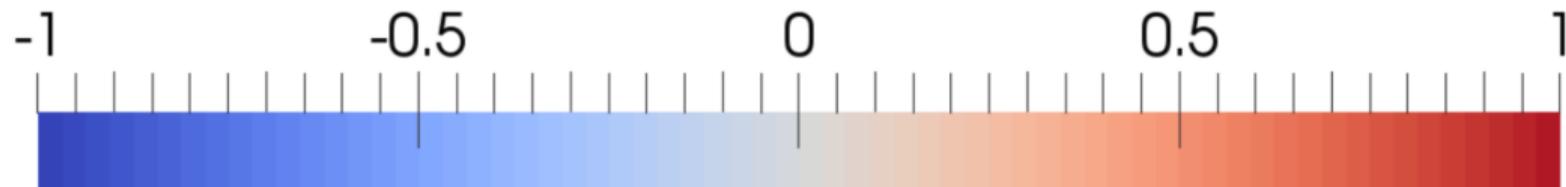


Abb. 5: Empfohlene Farbskala © Rechte siehe Anhang.

Anhang

Übersicht

Verzeichnis aller Abbildungen

Abb

Rechtsnachweise

©

Rechtliche Hinweise

§

Zitierweise dieses Dokuments

→

Verzeichnis aller Folien



1	Beispiel für einen Plot mit Error Bar	14
2	Problematik der Interpolation	15
3	Box-Plot	17
4	Rainbow Farbskala	18
5	Empfohlene Farbskala	19

Abb. 2 Quelle: <https://tex.stackexchange.com/questions/278210/draw-a-spline-with-pgfplots/278351>, Nutzung nach CC BY-SA <https://stackoverflow.com/help/licensing>.

Abb. 5 Quelle: Kenneth Moreland: Why we use bad color maps and what you can do about it. <http://www.kennethmoreland.com/color-advice/BadColorMaps.pdf>

Die hier angebotenen Inhalte unterliegen deutschem Urheberrecht. Inhalte Dritter werden unter Nennung der Rechtsgrundlage ihrer Nutzung und der geltenden Lizenzbestimmungen hier angeführt. Auf das Literaturverzeichnis wird verwiesen. Das **Zitatrecht** in dem für wissenschaftliche Werke üblichen Ausmaß wird beansprucht. Wenn Sie eine Urheberrechtsverletzung erkennen, so bitten wir um Hinweis an den auf der Titelseite genannten Autor und werden entsprechende Inhalte sofort entfernen oder fehlende Rechtsnennungen nachholen. Bei Produkt- und Firmennamen können Markenrechte Dritter bestehen. Verweise und Verlinkungen wurden zum Zeitpunkt des Setzens der Verweise überprüft; sie dienen der Information des Lesers. Der Autor macht sich die Inhalte, auch in der Form, wie sie zum Zeitpunkt des Setzens des Verweises vorlagen, nicht zu eigen und kann diese nicht laufend auf Veränderungen überprüfen.

Alle sonstigen, hier nicht angeführten Inhalte unterliegen dem Copyright des Autors, Prof. Dr. Clemens Cap, ©2020. Wenn Sie diese Inhalte nützlich finden, können Sie darauf verlinken oder sie zitieren. Jede weitere Verbreitung, Speicherung, Vervielfältigung oder sonstige Verwertung außerhalb der Grenzen des Urheberrechts bedarf der schriftlichen Zustimmung des Rechteinhabers. Dieses dient der Sicherung der Aktualität der Inhalte und soll dem Autor auch die Einhaltung urheberrechtlicher Einschränkungen wie beispielsweise **Par 60a UrhG** ermöglichen.

Die Bereitstellung der Inhalte erfolgt hier zur persönlichen Information des Lesers. Eine Haftung für mittelbare oder unmittelbare Schäden wird im maximal rechtlich zulässigen Ausmaß ausgeschlossen, mit Ausnahme von Vorsatz und grober Fahrlässigkeit. Eine Garantie für den Fortbestand dieses Informationsangebots wird nicht gegeben.

Die Anfertigung einer persönlichen Sicherungskopie für die private, nicht gewerbliche und nicht öffentliche Nutzung ist zulässig, sofern sie nicht von einer offensichtlich rechtswidrig hergestellten oder zugänglich gemachten Vorlage stammt.

Zitierweise dieses Dokuments

Wenn Sie Inhalte aus diesem Werk nutzen oder darauf verweisen wollen, zitieren Sie es bitte wie folgt:

Clemens H. Cap: Quantitatives wissenschaftliches Arbeiten. Electronic document.
<https://iuk.one/1012-1050> 5. 1. 2021.

Bibtex Information: <https://iuk.one/1012-1050.bib>

```
@misc{doc:1012-1050,  
  author      = {Clemens H. Cap},  
  title       = {Quantitatives wissenschaftliches Arbeiten},  
  year        = {2021},  
  month       = {1},  
  howpublished = {Electronic document},  
  url         = {https://iuk.one/1012-1050}  
}
```

Typographic Information:

Typeset on January 5, 2021

This is pdfTeX, Version 3.14159265-2.6-1.40.21 (TeX Live 2020) kpathsea version 6.3.2

This is pgf in version 3.1.5b

This is preamble-slides.tex myFormat©C.H.Cap

Verzeichnis aller Folien

- 1 Titelseite
- 2 Ziel
- 3 3 Arten von Skalen
- 4 Sonderformen metrischer Skalen
- 5 Rechenoperationen und Vergleiche auf Skalen
- 6 Kontinuierliche und Diskrete Skalen (1)
- 7 Kontinuierliche und Diskrete Skalen (2)
- 8 Beispiele für kontinuierliche Skalen (1)
- 9 Beispiele für kontinuierliche Skalen (2)
- 10 Typische Skalen-Fehler
- 11 Datenbereinigung
- 12 Checkliste: Wie gestaltet man eine Abbildung?
- 13 Wie plottet man Daten?

- 14 Fehlerbalken
- 15 Interpolation (1)
- 16 Interpolation (2)
- 17 Box Plot
- 18 Farbskalen (1)
- 19 Farbskalen (2)

Legende:

-  Fortsetzungsseite
-  Seite ohne Überschrift
-  Bildseite