

M24 Statistik 1: Sommersemester 2024

# Vorlesung 01: Intro

Prof. Matthias Guggenmos

Health and Medical University Potsdam

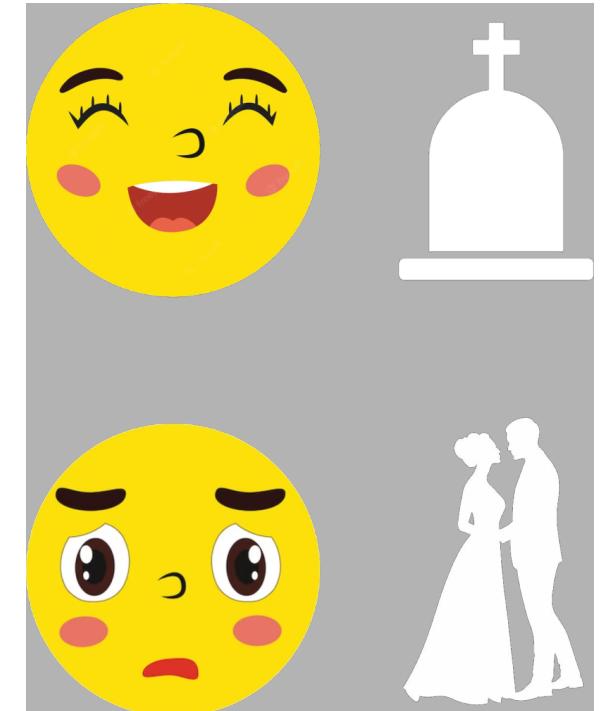


# Paradoxa

# Eine rätselhafte Erkrankung

Eine rätselhafte Erkrankung hat seit geraumer Zeit die Bundesrepublik Deutschland heimgesucht. Die Betroffenen zeigen ein paradoxes emotionales Verhalten: sie reagieren allem Anschein nach traurig auf fröhliche Ereignisse und fröhlich auf traurige Ereignisse – als wären zwei Hirndrähte vertauscht worden!

Aufgrund ihres Erscheinungsbilds wird die neuartige Krankheit Paradoxia genannt.



## Weiteres Mysterium: blaue Haare

Noch ist wenig bekannt über die Erkrankung. Eine weitere Auffälligkeit ist, dass viele der Betroffenen eine Neigung haben, ihre Haare blau zu färben. Warum das so ist, ist unklar, auch wenn gemutmaßt wird, dass hierfür ein viral gegangenes Tik-Tok-Video einer Betroffenen verantwortlich ist.



# Task Force

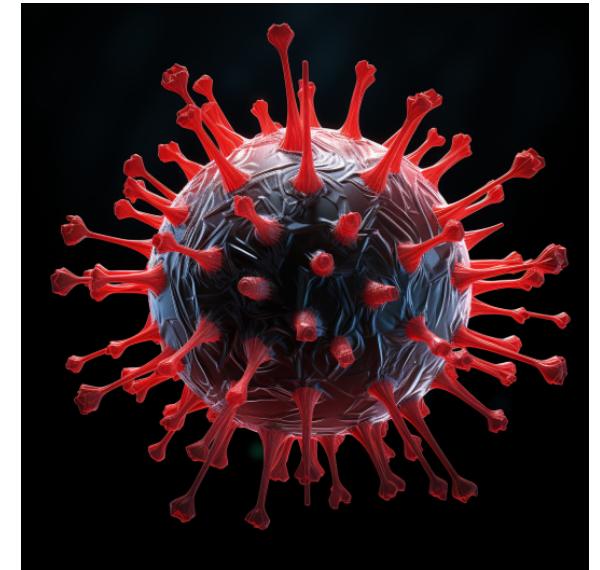
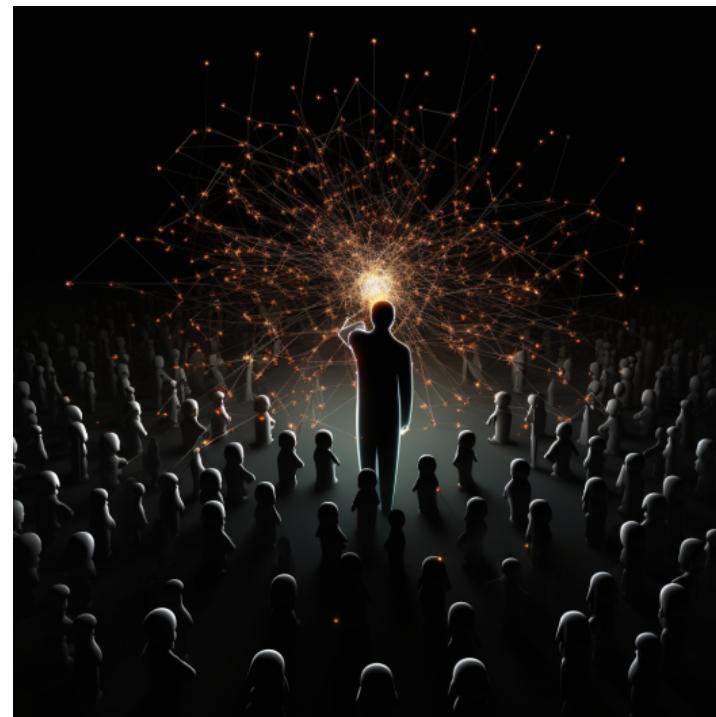
Durch die rapide Verbreitung der Erkrankung sind das Bundeskanzleramt und die Geheimdienste alarmiert. Schnell wird der Beschluss gefasst, eine Task Force bestehend aus Psycholog:innen, Mediziner:innen und Sozialforscher:innen ins Leben zu rufen, um die Erkrankung besser zu verstehen.



Sie sind Teil dieser Task Force im **Team Psychologie**.

# Eine erste Aufgabe: Hypothesenbildung

Ihre erste Aufgabe besteht darin, **Hypothesen** zu bilden, wie es zum Auftreten der Erkrankung gekommen sein könnte. Vermutet wird entweder ein **sozialpsychologisches Ansteckungsphänomen** oder eine **neuartige Viruserkrankung**, die das zentrale Nervensystem befällt.



# Aber zunächst: Formalitäten

# Anwesenheits- und Fehlzeitregelung

## § 8

### Anwesenheits- und Fehlzeitenregelung

- (1) Die nachfolgenden Anwesenheits-/Fehlzeitenregelungen gelten für alle Präsenzveranstaltungen. Für die Abwesenheit im Praktikum bestehen Regelungen in der jeweiligen Praktikumsordnung. Für die Abwesenheit/Fehlzeiten an Prüfungen/Prüfungstagen gelten vorrangig die Regelungen zum Rücktritt von Prüfungen nach der Rahmenprüfungsordnung oder den Studien- und Prüfungsordnungen der Studiengänge.
- (2) Die HMU ist eine Präsenzuniversität, damit besteht 100% Präsenzpflicht unter Beachtung der jeweiligen Studiengangmodelle (Vollzeitstudium, Teilzeitstudium). Pro Modul im Vollzeitmodell darf die Fehlzeit nicht mehr als 40% betragen, im Teilzeitmodell beträgt die maximale Fehlzeit 20% pro Modul. Im Studiengang Humanmedizin, im Bachelorstudiengang Psychologie und im Masterstudiengang Psychotherapie gilt ein Modul nur als regelmäßig besucht, wenn die Studierenden bei 80 % der Lehrveranstaltungen anwesend waren. Das gilt auch für virtuelle Präsenzlehrveranstaltungen. Sollte diese Fehlzeit von Studierenden überschritten werden, entscheidet der Lehrende über die Zulassung zur Prüfung. In

Bildnachweis<sup>1</sup>

# Nützliche Links

Statistik für Psychologie	<a href="https://www.statistikpsychologie.de/blog/">https://www.statistikpsychologie.de/blog/</a>	Einige nützliche Blogeinträge zu statistischen Themen der Psychologie
DATAtab	<a href="https://datatab.de/tutorial/">https://datatab.de/tutorial/</a>	Online-Kalkulator mit einigen guten Tutorials
StatistikGuru	<a href="https://statistikguru.de/">https://statistikguru.de/</a>	Fokus auf SPSS und R, aber viele Tutorials in allgemeinem Stil geschrieben
Statistics Supreme with JASP	<a href="https://www.youtube.com/@StatisticsSupreme">https://www.youtube.com/@StatisticsSupreme</a>	YouTube-Kanal zu JASP des Kollegen Thomas Langkamp (MSH)
Statistik 1 Vorlesungen auf YouTube	<a href="https://www.youtube.com/playlist?list=PLU-U4gWcFfPxmKviyopGSmOY2EaKLVYxF">https://www.youtube.com/playlist?list=PLU-U4gWcFfPxmKviyopGSmOY2EaKLVYxF</a>	Prof. Jonas Obleser (Uni Lübeck)

# Vorlesungsfolien

Trainex

The screenshot shows the HMU Trainex interface. On the left, there's a sidebar with links for Anwesenheiten (Archive), E-Learning, Glossar, Hausaufgaben, Laufwerk, and Bibliothek. The main content area displays course information for M24 Statistik I, including a link to Feinplanung (PDF, 61 KB) and a link to Einführung in den Bachelor Psychologie (PDF).

Format: pdf

[trainex25.de/hmu-trainex](http://trainex25.de/hmu-trainex)

Onlineversion

The screenshot shows a web browser window displaying the online version of the lecture slides. The title is "Vorlesung 01: Intro - Chromium". The page content includes the title "Vorlesung 01: Intro", author "Prof. Matthias Guggenmos", institution "Health and Medical University Potsdam", date "10.10.2023", and a small logo. The URL in the address bar is "m-guggenmos.github.io/01\_intro/\_slides/index.html#/".

Format: html & pdf

[m-guggenmos.github.io](https://m-guggenmos.github.io) oder  
[cutt.ly/risch](https://cutt.ly/risch)

# Klausur

- In der vorlesungsfreien Zeit nach Semesterende findet eine **schriftliche Klausur** zu Statistik 1 statt.
- **Klausurrelevant** sind dabei alle **Inhalte der Vorlesungsfolien**, falls nicht extra gekennzeichnet.
- **Klausurrelevant** sind auch die **Übungen**, die Sie im **Seminar** durchführen.



# Literatur



Jürgen Bortz & Christof Schuster. *Statistik für Human- und Sozialwissenschaftler*. Springer, 7. Auflage (2010).

Standardwerk. Umfassend, gute Erklärungen, viele Übungen.



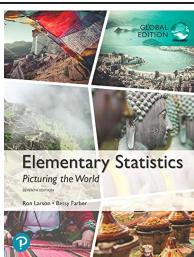
Thomas Schäfer. *Methodenlehre und Statistik*. Springer, 1. Auflage (2016).

Nah an der Vorlesung, Inhalte auf das Wesentliche heruntergebrochen.



Michael Eid, Mario Gollwitzer & Manfred Schmitt. *Statistik und Forschungsmethoden*. Beltz, 5. Auflage (2017).

Umfassend, vergleichsweise mathematisch, eher für Fortgeschrittene und weitergehend Interessierte. Enthält nützliche Übungen.

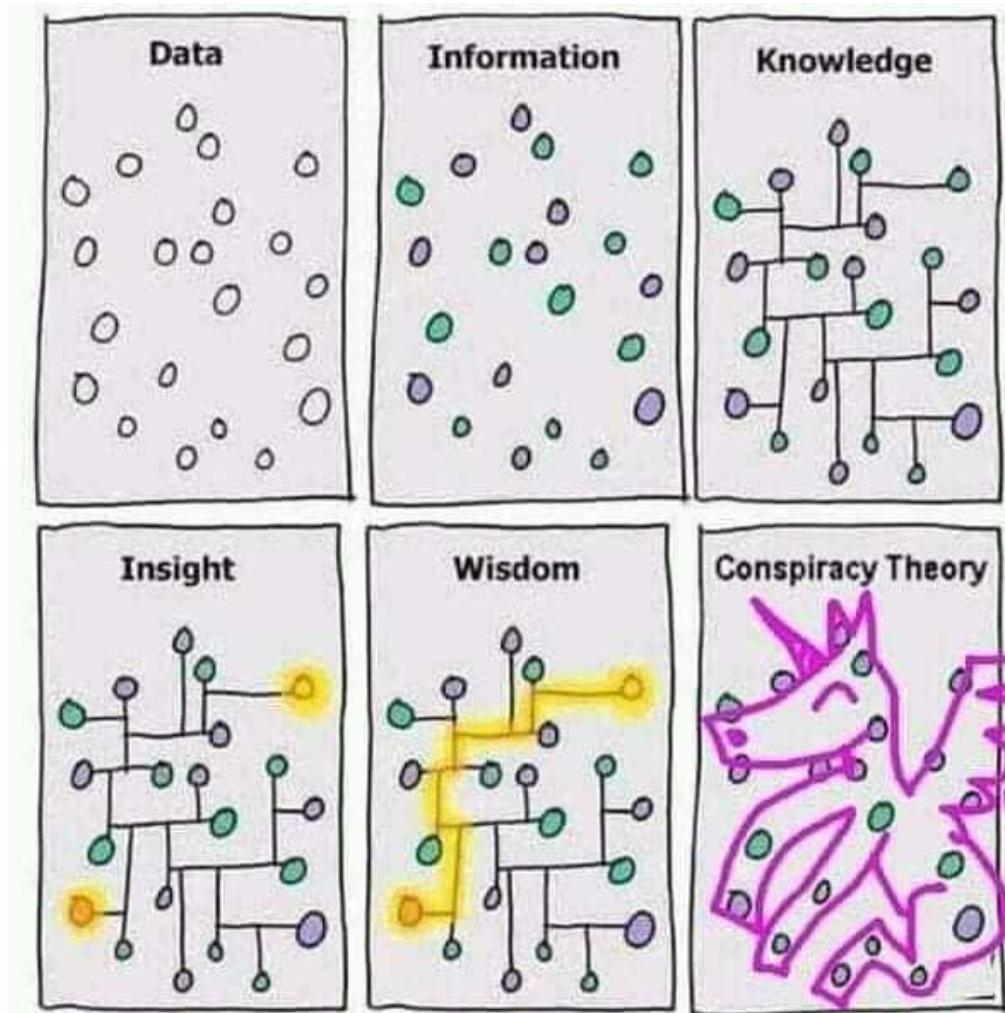


Ron Larson & Betsy Farber. *Elementary Statistics: Picturing the World*. Pearson, 7. Auflage (2019).

Unterhaltsamer intuitiver Stil, kurz & bündig, viele Übungen, auf Englisch!

# Was ist Statistik?

# Daten, Informationen, Verschwörungstheorien



Bildnachweis<sup>2</sup>

# Was ist Statistik?

Das Wort Statistik stammt von lateinisch *statisticum* „den Staat betreffend“ und italienisch *statista* Staatsmann oder Politiker [...]

Quelle: Wikipedia

In diesem Sinne beschrieb der Ausdruck Statistik ursprünglich die Lehre von den Daten über den Staat. Heute ist der Begriff wesentlich breiter gefasst:

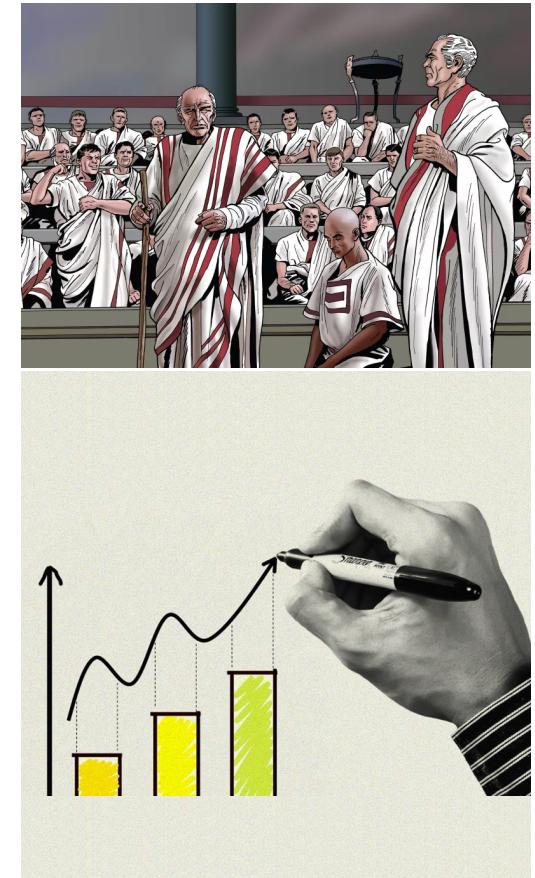
---

**Statistik** ist ein angewandter Zweig der Mathematik, der sich mit der **Erhebung, Analyse und Interpretation** empirischer Daten befasst.

---

## Beispiele

<b>Erhebung</b>	Wie groß sollte meine Stichprobe sein?
<b>Analyse</b>	Was ist der Mittelwert meiner Variable?
<b>Interpretation</b>	Unterscheiden sich zwei Gruppen systematisch?



Bildnachweis<sup>3 4</sup>

# Wozu Statistik?

Auch ohne Statistik sind Menschen sehr gut darin, Muster oder Zusammenhänge in ihrer Umwelt zu erkennen. Man könnte diese Fähigkeit als **intuitive Empirie** bezeichnen. Wozu also Statistik?

## Probleme intuitiver Empirie (kleiner Ausschnitt)

- Kognitive Biases (z.B. Konfirmationsbias)
- Schwächen beim probabilistischen Denken (Stichwort *Lotto*)
- Übergeneralisierung / Tendenz zur vorschnellen Verallgemeinerung
- Übersensitivität: Muster erkennen, wo keine sind
- Vermischung mit eigenen Interessen (z.B. Politik)
- Begrenztes (Arbeits)Gedächtnis: Unfähigkeit, große Datenmengen im Kopf zu behalten und zu verrechnen



Bildnachweis<sup>5</sup>

# Beispiel Prävalenzfehler

Ein neu entwickelter Coronatest wird damit beworben, dass er anhand einer Speichelprobe in 100% der Fälle eine bestehende Coronaerkrankung erkennt. Im Kleingedruckten findet sich der Hinweis, dass die Spezifität 90% beträgt, d.h. der Test weist in 10% der Fälle eine gesunde Person fälschlicherweise als krank aus.

Nehmen Sie an, die Prävalenz von Corona ist 1%, d.h. eine von hundert Personen ist mit Corona infiziert.



**Schätzen Sie:** Angenommen Ihr Test ist positiv, in wie viel Prozent der Fälle sind Sie dann auch tatsächlich positiv?

Join at [menti.com](https://menti.com) use code 2855 7996

 **Mentimeter**

The code lets your audience join the presentation and expires in 2 days.



## Schätzfrage zu Corona-Tests

Teilnahme über Zahlencode oder QR-Code

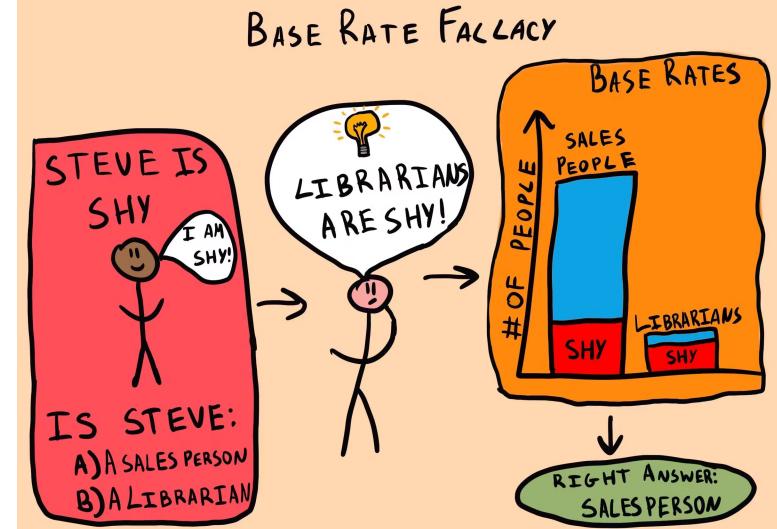
<https://www.mentimeter.com/>

# Beispiel Prävalenzfehler

Auflösung:

- von 100 Personen ist im Schnitt 1 Person krank. Diese wird durch den Test auf jeden Fall korrekt als krank erkannt.
- 99 Personen sind nicht erkrankt, und bei 10% dieser Personen gibt der Test ein falsch-positives Ergebnis ( $\approx 10$  Personen).
- D.h. bei 100 Personen haben  $\approx 11$  Personen ein positives Ergebnis, aber bei 10 von 11 Personen (91%) ist dieses Ergebnis falsch!

Unsere empirische Intuition versagt in diesem Fall häufig. Im konkreten Beispiel spricht man auch vom Prävalenzfehler oder der Base rate fallacy.



Weiteres Beispiel: ist die Häufigkeit/Prävalenz von "Sales people" deutlich größer als die der "Librarians", so ist die Wahrscheinlichkeit groß, dass eine schüchterne Person mit großer Wahrscheinlichkeit eine "Sales person" – selbst dann, wenn Sales people mehrheitlich eher nicht schüchtern sind.<sup>6</sup>

# Beispiel Meditonsin

Meditonsin ist das meistverkaufte homöopathische Erkältungsmittel in Deutschland, mit einem Jahresumsatz von deutlich über zwanzig Millionen Euro.

Jedenfalls entwickelte ein Chemieprofessor für Medice eine Apparatur, die alle Zutaten von Meditonsin bis heute in 250-Liter-Fässern vermischt und am Ende zehnmal kräftig in Richtung Erdmittelpunkt schüttelt.

Meditonsin enthält laut Beipackzettel Tollkirsche und Eisenhut in D5-Auflösung [0,00001%] sowie Quecksilber in D8-Auflösung [0,00000001%].

Quelle: Sueddetsche Zeitung<sup>7</sup>

## DAZ.online

Unabhängige pharmazeutische Informationen für Wissenschaft und Praxis der Deutschen Apotheker Zeitung

### RECHTSKRÄFTIGES URTEIL

#### Meditonsin-Werbung mit Apotheken-Studie unzulässig

BERLIN - 03.05.2023, 14:45 UHR



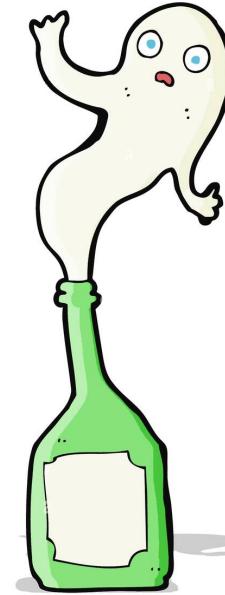
Meditonsin-Hersteller Medice darf für seine homöopathischen Tropfen nicht mehr damit werben, dass in einer „apothekenbasierten Beobachtungsstudie“ die gute Wirksamkeit und Verträglichkeit des Mittels nachgewiesen wurde. Ein entsprechendes Urteil des Landgerichts Dortmund aus dem vergangenen Herbst ist jetzt rechtskräftig geworden. Geklagt hatte die Verbraucherzentrale Nordrhein-Westfalen.

Laut einem Gerichtsurteil aus dem Mai 2023 darf Meditonsin nicht mehr mit einer vorgeblich wissenschaftlich bewiesenen Wirksamkeit werben.<sup>8</sup>

# Beispiel Meditonsin

Woher kommt das große Vertrauen in ein Medikament, bei dem eine Flasche im Schnitt kaum ein Molekül der vorgeblichen Wirksubstanzen enthält?

Warum entsteht bei vielen Menschen subjektiv der Eindruck, Medikamente wie Meditonsin würden wirken?

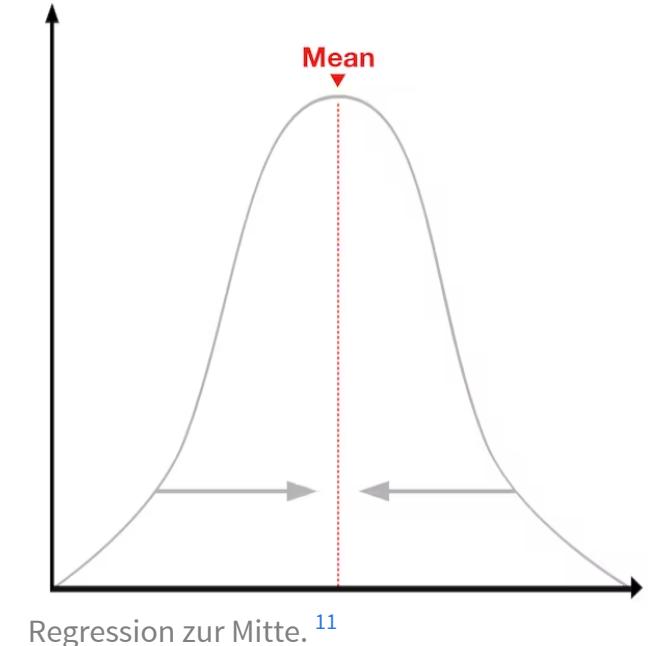


Der Geist in der Flasche.<sup>9</sup>

# Beispiel Meditonsin

Eine Erklärung ist mangelndes Bewusstsein für einen Effekt, der auch als **Regression zur Mitte** bezeichnet wird.

- Meditonsin wird vermutlich häufig eingenommen, wenn eine Erkältung auf ihrem Höhepunkt ist.
- *In aller Regel verbessert* sich der Gesundheitszustand nach Erkältungen auch ohne Medikation und pendelt sich wieder auf einen Normalzustand ein (dies ist die “Regression” zur “Mitte”).
- Dadurch entsteht eine subjektiv wahrgenommener Zusammenhang Einnahme Meditonsin → Heilung



**Statistische Methoden** ermöglichen hier in Kombination mit einem adäquaten Studiendesign (Interventions- und Placebogruppe) eine objektivere Einschätzung.

Empfehlung: Quarks Science Podcast “Meditonsin, Umckaloabo, Grippostad: Die Akte Erkältungsmittel”<sup>10</sup>

# Deskriptive Statistik und Inferenzstatistik

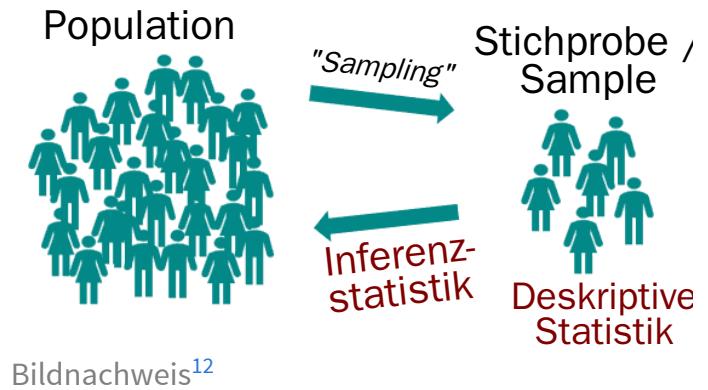
In Statistik 1 und 2 werden wir Methoden aus der **deskriptiven Statistik** und **Inferenzstatistik** behandeln.

Das Ziel der **deskriptiven Statistik** ist die Beschreibung und Zusammenfassung von Daten innerhalb einer Stichprobe.

Beispielfragen: Wie hoch ist der Mittelwert einer Variable in der Stichprobe? Wie hoch ist die Korrelation zwischen Variable X und Variable Y in der Stichprobe?

Das Ziel der **Inferenzstatistik** sind Schlussfolgerungen über eine Population auf der Grundlage von Stichproben.

Beispielfragen: Wie groß ist die durchschnittliche Körpergröße der Bevölkerung, basierend auf einer Stichprobe von 500 Personen? Gibt es einen signifikanten Unterschied im durchschnittlichen Einkommen zwischen zwei verschiedenen Berufsgruppen?

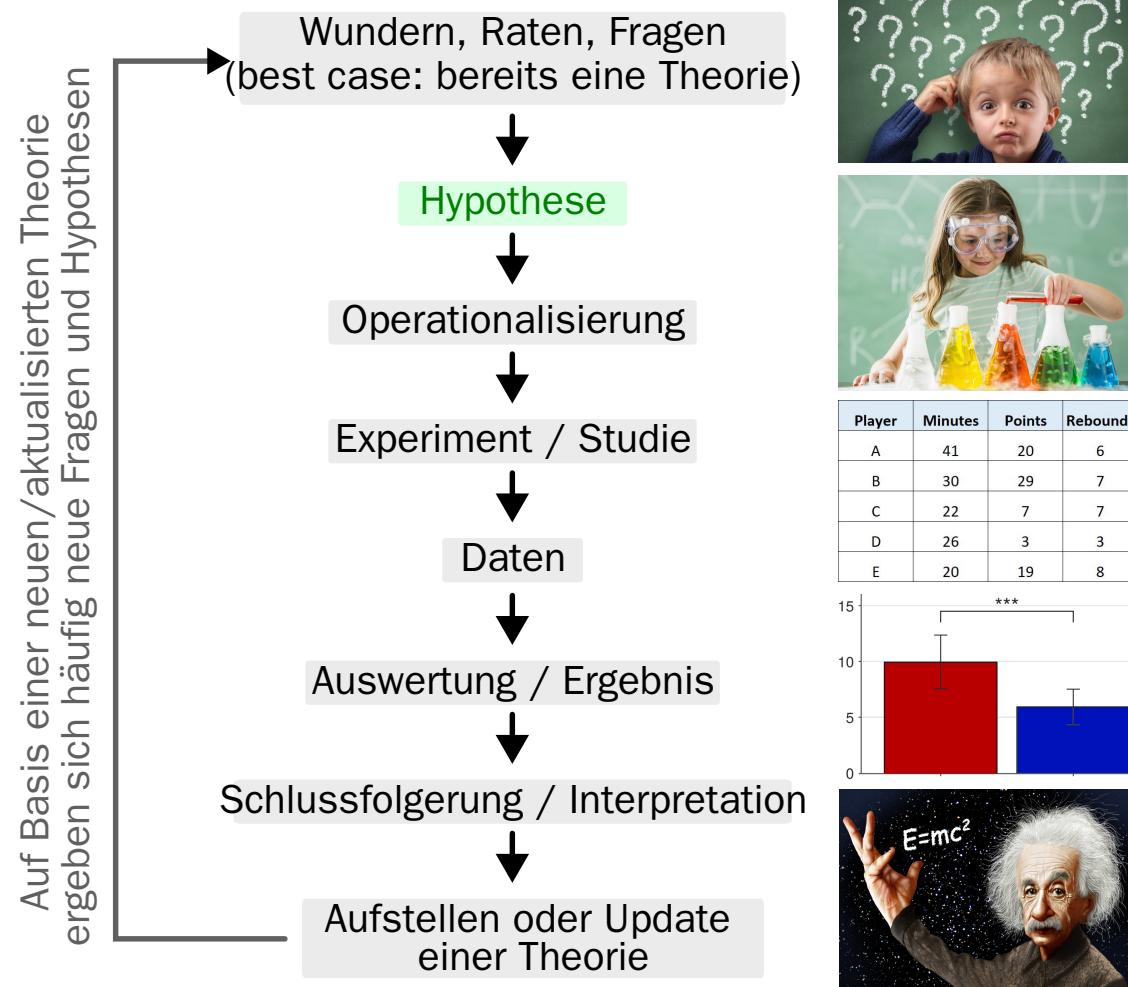


Bildnachweis<sup>12</sup>

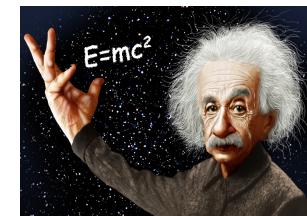
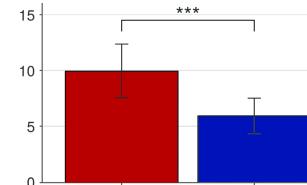
Ein alternativer Ausdruck für Inferenzstatistik ist Induktive Statistik.

# Hypothesen

# Der Forschungsprozess



Player	Minutes	Points	Rebounds
A	41	20	6
B	30	29	7
C	22	7	7
D	26	3	3
E	20	19	8

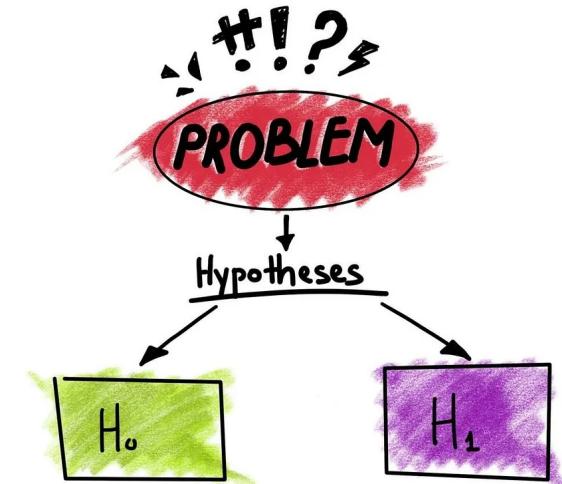


# Was ist eine Hypothese?

**Hypothesen:** in Form einer **logischen Aussage** formulierte **Annahme** zur Erklärung **empirischer Phänomene**.

## Beispiele

- Es gibt einen positiven Zusammenhang zwischen der Feinstaubbelastung einer Stadt und der Prävalenz von Lungenerkrankungen.
- Eine vorherige Impfung mit dem Corona-Impfstoff verringert die Symptomschwere im Fall einer Infektion.



Bildnachweis<sup>13</sup>

# Gute Hypothesen, schlechte Hypothesen

Eine wissenschaftliche Hypothese muss drei Grundvoraussetzungen erfüllen: **Widerspruchsfreiheit, Widerlegbarkeit, Operationalisierbarkeit.**

## Beispiele

- Alle Lebewesen benötigen Sauerstoff, aber manche Lebenwesen sind anaerob [widerspruchsfrei, widerlegbar, operationalisierbar]
- Könnten Babies sprechen, würden sie hauptsächlich ihre Gefühle zum Ausdruck bringen [widerspruchsfrei, operationalisierbar, widerlegbar]
- Gott ist eine notwendige Voraussetzung für Leben [widerspruchsfrei, operationalisierbar, widerlegbar].

**Beachte:** Hypothese, die nicht widerspruchsfrei oder nicht operationalisierbar sind, sind auch nicht widerlegbar (die umgekehrte Aussage gilt nicht in allen Fällen).

**Operationalisierung:** Messbarmachung eines theoretischen Konstrukt (Schwerkraft, Intelligenz, psychische Resilienz, ...) durch ein geeignetes Messinstrument (Gravimeter, Intelligenztest, Fragebogen, ...)

## Eine gute wissenschaftliche Hypothese erfüllt darüber hinaus weitere Kriterien:

<b>Prädiktivität</b>	<p><i>Hypothesen sollten präzise Vorhersagen ermöglichen</i></p> <p><u>Negativbeispiel</u>: Das Selbst ist in ein Es, Ich, Über-Ich aufgeteilt.</p> <p><u>Positivbeispiel</u>: Eine tägliche morgendliche Meditationsübung verringert das subjektive und objektive Stresslevel.</p>
<b>Spezifität</b>	<p><i>Hypothesen sollten spezifisch und konkret sein</i></p> <p><u>Negativbeispiel</u>: Frühe Kindheitstraumata wirken sich nachteilig auf spätere Beziehungen aus.</p> <p><u>Positivbeispiel</u>: Kindheitstraumata in einem vorsprachlichen Alter, verursacht durch wichtige Bezugspersonen, erhöhen die Wahrscheinlichkeit von Bindungsangst in engen romantischen Beziehungen.</p>
<b>Sparsamkeit</b>	<p><i>Hypothesen sollten keine unnötige Komplexität beinhalten</i></p> <p><u>Negativbeispiel</u>: Die Coronapandemie verstärkt das Gefühl der Vereinsamung in älteren Bevölkerungsschichten, falls diese an einem Dienstag befragt wurden.</p> <p><u>Positivbeispiel</u>: Die Coronapandemie verstärkt das Gefühl der Vereinsamung in älteren Bevölkerungsschichten, <del>falls diese an einem Dienstag befragt wurden</del>.</p>
<b>Relevanz</b>	<p><i>Hypothesen sollten von theoretischer oder gesellschaftlicher Relevanz sein</i></p> <p><u>Negativbeispiel</u>: Menschen mit roten Autos bevorzugen rote Gemüsesorten.</p> <p><u>Positivbeispiel</u>: Nach persönlichen Erfolgserlebnissen neigen Menschen zu riskantem Fahrverhalten.</p>

# Gerichtete und ungerichtete Hypothesen

**Ungerichtete Hypothesen** geben keine Richtung des Effektes an:

- Frauen und Männer haben unterschiedliche viele soziale Kontakte
- Die Punktzahl im Abitur hängt zusammen mit der späteren Lebenszufriedenheit

**Gerichtete Hypothesen** geben die Richtung des Effektes an:

- Frauen haben *mehr* soziale Kontakte als Männer
- Eine höhere Punktzahl im Abitur korreliert *negativ* mit der späteren Lebenszufriedenheit



Bildnachweis<sup>14</sup>

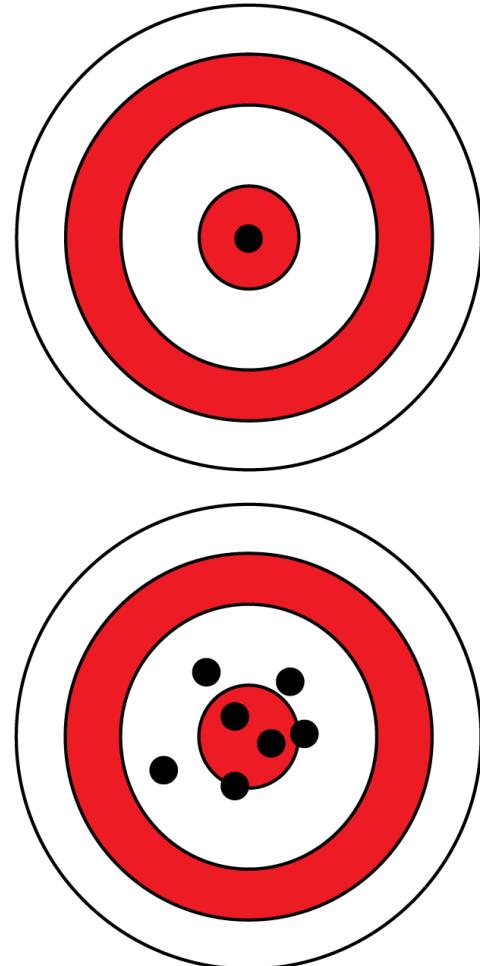
# Probabilistische Hypothesen

In der Psychologie sind Hypothesen in aller Regel nicht **deterministisch**, sondern **probabilistisch**, d.h. sie machen Aussagen über Wahrscheinlichkeiten und Mittelwerte.

**Beispiel deterministische Hypothese:** Menschen mit Alkoholabhängigkeit entwickeln eine Depression.

**Beispiel probabilistische Hypothese:** Menschen mit Alkoholabhängigkeit haben *im Mittel* eine höhere Ausprägung depressiver Symptome.

- Psychologische Phänomene sind in der Regel von einer Vielzahl von Faktoren beeinflusst, die zum Großteil nicht bekannt sind (biologisch-z.B. Blutdruck, psychologisch-z.B. Stimmung, situativ-z.B. Versuchslabor versus Realität).
- Diese Faktoren erhöhen die Variabilität von Beobachtungen und führen dazu, dass die meisten psychologischen Phänomene nur durch Mittelwertbildung über eine Reihe von Messwiederholungen sichtbar werden.



Bildnachweis<sup>15</sup>



Ok, Task Force: wie lauten eure (Arbeits-)Hypothesen zu den Hintergründen von Paradoxa?

**Hypothese 1:** Die Blaufärbung der Haare ist ein reines Tik-Tok-Phänomen. *Betroffene sollten daher mehr Zeit auf der Social-Media-Plattform TikTok verbringen als Nicht-Betroffene.*

**Hypothese 2:** Paradoxa wird durch eine virale Infektion verursacht. *Die Blutproben von Betroffenen sollten daher eine höhere Konzentration von Entzündungsmarkern aufweisen als von Nicht-Betroffenen.*

# Fußnoten

1. Allgemeine Universitätsordnung der HMU
2. <https://languagelog.ldc.upenn.edu/nll/?p=52581>
3. <https://www.analyticsinsight.net/importance-of-statistics-how-is-statistics-related-to-data-science/>
4. <https://www.livescience.com/roman-republic>
5. <https://www.arte.tv/de/videos/093029-022-A/flick-flack/>
6. <https://thedecisionlab.com/biases/base-rate-fallacy>
7. <https://sz-magazin.sueddeutsche.de/gesundheit/meditonsin-wirkung-erkaeltung-globuli-81941>
8. <https://www.astronomy.com/science/in-praise-of-nothing/>
9. <https://www.alamy.de/stockfoto-cartoon-geist-in-der-flasche-72046687.html>
10. <https://www.quarks.de/podcast/quarks-science-cops-folge-48-erkaeltungsmittel-wissenschaft-oder-kruemeltee/>
11. <https://theconversation.com/regression-to-the-mean-or-why-perfection-rarely-lasts-74694>
12. <https://databab.net/tutorial/descriptive-inferential-statistics>
13. <https://pub.towardsai.net/hypotheses-testing-with-scipy-b5ba86430d74>
14. <https://databab.de/tutorial/hypothesen>
15. <https://www.toolsgroup.com/blog/what-is-probabilistic-forecasting/>