

# Statistik - Kapitel 01

Karsten Keßler

## Inhaltsverzeichnis

<b>1</b>	<b>Einführung in die Statistik</b>	<b>2</b>
1.1	Was bedeutet „Statistik“?	2
1.1.1	Statistik als Datenmaterial	2
1.1.2	Statistik als Methode	2
1.1.3	Statistik als Wissenschaft	2
1.2	Warum brauchen wir Statistik?	3
1.2.1	Warum Informatiker Statistik brauchen	3
1.2.2	Das Problem der Variabilität	3
1.3	Grundbegriffe	4
1.3.1	Statistische Einheit (Merkmalsträger)	4
1.3.2	Merkmal und Merkmalsausprägung	4
1.3.3	Grundgesamtheit und Stichprobe	5
1.3.4	Bias in der Informatik	5
1.4	Messniveaus (Skalentypen)	6
1.4.1	Nominalskala	6
1.4.2	Ordinalskala	6
1.4.3	Metrische Skala	7
1.4.4	Typische Fallen in der Informatik	7
1.4.5	Übersicht: Erlaubte Operationen	8
1.5	Deskriptive vs. induktive Statistik	8
1.5.1	Deskriptive Statistik	8
1.5.2	Induktive Statistik (schließende Statistik)	8
1.5.3	Vergleich: Deskriptiv vs. Induktiv	8
1.5.4	Praxisbeispiel: A/B-Test	9
1.6	Typischer Statistik-Workflow	9
1.6.1	Workflow am IT-Beispiel: SLA-Prüfung	10
1.6.2	Typische Fehler im Workflow	10
1.7	Anwendungsbeispiele aus der Informatik	10
1.7.1	Beispiel 1: Performance-Monitoring	10
1.7.2	Beispiel 2: A/B-Test für UI-Änderung	11
1.7.3	Beispiel 3: Log-Analyse für Anomalie-Erkennung	11
1.8	Zusammenfassung	12
1.8.1	Zusammenfassung für Informatiker	12
1.8.2	Messniveau-Cheatsheet für IT-Merkmale	12
1.8.3	Wichtige Prinzipien	12

Stand: 17. Januar 2026

# 1 Einführung in die Statistik

## Leitfrage:

Wie kommt man von den „rohen Zahlen“ zu begründeten Aussagen und Entscheidungen?

## Lernziele:

- (i) grundlegende Begriffe der Statistik sauber verwenden,
  - (ii) statistische Fragestellungen präzise formulieren und
  - (iii) den Unterschied zwischen Deskription und Induktion erklären.
- 

## 1.1 Was bedeutet „Statistik“?

Der Begriff **Statistik** wird in der Praxis in drei typischen Bedeutungen verwendet:

### 1.1.1 Statistik als Datenmaterial

Veröffentlichte Zahlen, Tabellen oder Kennzahlen

(z. B. Umsatzstatistik, Arbeitslosenstatistik, Wahlstatistik).

→ Statistik als **Ergebnis** einer Datenerhebung.

#### Informatik-Beispiele:

- Server-Uptime-Statistiken (99,9% Verfügbarkeit)
- GitHub-Repository-Metriken (Stars, Forks, Contributors)
- App-Store-Bewertungsstatistiken
- Website-Traffic-Reports (Visits, Bounce-Rate)
- Fehlerstatistiken aus Bug-Tracking-Systemen

### 1.1.2 Statistik als Methode

Verfahren zur **Erhebung, Aufbereitung, Darstellung** und **Analyse** von Daten:

Häufigkeitstabellen, Diagramme, Mittelwert/Varianz, Modelle, Tests.

#### Informatik-Beispiele:

- Log-Analyse und Aggregation (ELK-Stack, Splunk)
- A/B-Testing-Frameworks
- Performance-Monitoring (Grafana, Prometheus, Datadog)
- Machine-Learning-Evaluationsmetriken (Accuracy, F1-Score)
- Code-Qualitätsmetriken (SonarQube)

### 1.1.3 Statistik als Wissenschaft

Mathematische Disziplin, die Modelle für **Zufall und Unsicherheit** entwickelt und untersucht, wie zuverlässig Schlussfolgerungen aus Stichproben sind.

#### Informatik-Bezüge:

- Wahrscheinlichkeitstheorie für Kryptographie
- Stochastische Prozesse für Netzwerk- und Warteschlangenanalyse
- Bayessche Inferenz für Machine Learning
- Statistische Lerntheorie (PAC-Learning, VC-Dimension)
- Information Theory (Shannon-Entropie)



### Merksatz

In dieser Vorlesung steht Statistik vor allem als **Methode** und **Wissenschaft** im Vordergrund.

## 1.2 Warum brauchen wir Statistik?

In vielen realen Situationen sind Ergebnisse **nicht deterministisch** und Daten **unvollständig**:

- In der Qualitätskontrolle wird nicht jedes Teil geprüft.
- In Umfragen werden nicht alle Personen befragt.
- In IT-Systemen schwanken Antwortzeiten oder Ausfallraten.
- In der Wirtschaft variieren Nachfrage, Preise und Lieferzeiten.

### 1.2.1 Warum Informatiker Statistik brauchen

Anwendungsbereich	Typische statistische Fragen
<b>Software-Testing</b>	Wie viele Tests reichen für ausreichende Abdeckung?
<b>Performance-Optimierung</b>	Ist die neue Version wirklich schneller?
<b>Machine Learning</b>	Wie gut generalisiert mein Modell auf neue Daten?
<b>Security/Anomalie-Erkennung</b>	Ist dieses Login-Verhalten verdächtig?
<b>Kapazitätsplanung</b>	Wie viele Server brauchen wir für Spitzenlasten?
<b>A/B-Testing</b>	Hat die UI-Änderung einen signifikanten Effekt?
<b>Datenqualität</b>	Wie gehen wir mit fehlenden oder fehlerhaften Daten um?
<b>SLA-Monitoring</b>	Erfüllen wir die vereinbarten Service Level?

### 1.2.2 Das Problem der Variabilität

Selbst bei identischen Bedingungen variieren IT-Metriken:

**Beispiel:** Dieselbe Datenbankabfrage, 5× ausgeführt:

Messung	Zeit (ms)
1	12,3
2	14,1
3	11,8
4	45,2
5	13,5

**Fragen:**

- Was ist die „typische“ Antwortzeit?

- Ist Messung 4 ein Problem oder normal?
- Wie zuverlässig ist ein einzelner Messwert?

→ Statistik liefert die Werkzeuge, um solche Fragen systematisch zu beantworten.

Statistik liefert Werkzeuge, um

- Daten verständlich zu strukturieren,
- Zufall von Struktur zu trennen,
- Unsicherheit zu quantifizieren,
- Entscheidungen nachvollziehbar zu begründen.



### Wichtig:

Statistik liefert selten Gewissheit, sondern **begründete Aussagen mit Fehlerwahrscheinlichkeit**.

## 1.3 Grundbegriffe

### 1.3.1 Statistische Einheit (Merkmalsträger)

Die **statistische Einheit** ist das Objekt, über das Daten erhoben werden (kleinste Untersuchungseinheit).

**Allgemeine Beispiele:** Person, Bestellung, Produkt, Bauteil, Messung

**Informatik-Beispiele:**

Kontext	Statistische Einheit
Web-Analytics	Ein Seitenaufruf (Page View)
API-Monitoring	Ein einzelner HTTP-Request
Datenbank-Analyse	Eine Query / Transaktion
User-Research	Ein Nutzer / Eine Session
Log-Analyse	Eine Log-Zeile / Ein Event
Software-Testing	Ein Testfall / Ein Test-Run
Code-Qualität	Eine Datei / Funktion / Klasse
DevOps	Ein Deployment / Ein Build

### 1.3.2 Merkmal und Merkmalsausprägung

- **Merkmal:** Eigenschaft der Einheit (z. B. *Lieferzeit*)
- **Merkmalsausprägung:** konkret beobachteter Wert (z. B. *3,2 Tage*)



### Mini-Beispiel:

Einheit = Lieferung · Merkmal = Lieferzeit · Ausprägung = 3,2 Tage

**Informatik-Beispiel:**

Einheit	Merkmal	Ausprägung
HTTP-Request	Response-Zeit	142 ms
HTTP-Request	Statuscode	200
HTTP-Request	HTTP-Methode	GET
Server	CPU-Auslastung	73,5 %
Bug-Ticket	Priorität	High
Code-Commit	Lines of Code	247

### 1.3.3 Grundgesamtheit und Stichprobe

- **Grundgesamtheit:** alle relevanten Einheiten, über die eine Aussage getroffen werden soll
- **Stichprobe:** tatsächlich beobachtete Teilmenge

#### Beispiel:

Grundgesamtheit = alle Lieferungen eines Jahres

Stichprobe = 120 zufällig ausgewählte Lieferungen

#### Informatik-Beispiele:

Grundgesamtheit	Stichprobe
Alle Requests eines Tages (10 Mio.)	10.000 zufällig geloggte Requests
Alle User einer Plattform (1 Mio.)	500 User für eine Usability-Studie
Alle möglichen Eingaben eines Programms	Ausgewählte Testfälle
Alle Commits eines Projekts	Die letzten 100 Commits
Alle potenziellen Kunden	A/B-Test mit 5% der Besucher



#### Repräsentativität:

Die Stichprobe soll die Grundgesamtheit möglichst gut abbilden.

Verzerrungen (Bias) führen zu falschen Schlussfolgerungen.

### 1.3.4 Bias in der Informatik

#### Typische Verzerrungen bei IT-Daten:

Bias-Art	Beschreibung	Beispiel
<b>Survivorship Bias</b>	Nur „überlebende“ Daten werden analysiert	Nur erfolgreiche Requests im Log
<b>Selection Bias</b>	Systematisch verzerrte Auswahl	Nur Power-User geben Feedback
<b>Temporal Bias</b>	Zeitliche Verzerrung	Daten nur von Werktagen, nicht Wochenenden
<b>Sampling Bias</b>	Nicht-repräsentative Stichprobe	Requests nur aus einem Rechenzentrum

Bias-Art	Beschreibung	Beispiel
<b>Measurement Bias</b>	Messung beeinflusst Ergebnis	Profiling verlangsamt die Anwendung

### Beispiel Survivorship Bias:

Ein Monitoring-System loggt nur Requests, die erfolgreich beantwortet wurden. Requests mit Timeout werden nicht erfasst.

→ Die gemessene durchschnittliche Response-Zeit ist **systematisch zu niedrig**, weil die langsamsten Requests (die mit Timeout) fehlen.

## 1.4 Messniveaus (Skalentypen)

Das **Messniveau** bestimmt, welche statistischen Auswertungen sinnvoll und zulässig sind.

### 1.4.1 Nominalskala

- Kategorien ohne natürliche Ordnung
- Erlaubte Operationen: Gleichheit (=, )
- Beispiele: Farbe, Land, Produkttyp
- Auswertungen: Häufigkeiten, Modus, Balkendiagramme

#### Informatik-Beispiele:

- HTTP-Methode (GET, POST, PUT, DELETE, PATCH)
- Betriebssystem (Windows, macOS, Linux, Android, iOS)
- Programmiersprache (Python, Java, C++, JavaScript)
- Browser (Chrome, Firefox, Safari, Edge)
- Error-Typ (Timeout, NotFound, ServerError, AuthError)
- Serverstandort (Frankfurt, Dublin, Virginia, Singapore)

### 1.4.2 Ordinalskala

- Geordnete Kategorien ohne interpretierbare Abstände
- Erlaubte Operationen: Gleichheit + Ordnung (<, >, , )
- Beispiele: Schulnoten, Zufriedenheit (1–5), Rankings
- Auswertungen: Median, Quartile, Rangmaße  
(Mittelwert nur mit Vorsicht!)

#### Informatik-Beispiele:

- Bug-Priorität (Critical, High, Medium, Low)
- User-Rating (1–5 Sterne)
- Log-Level (DEBUG, INFO, WARNING, ERROR, CRITICAL)
- Story Points / T-Shirt-Sizes (XS, S, M, L, XL)
- Service-Tier (Free, Basic, Pro, Enterprise)
- Reifegrad (Alpha, Beta, Stable, LTS)

**Achtung:** Der Abstand zwischen „3 Sterne“ und „4 Sterne“ ist nicht notwendigerweise gleich dem Abstand zwischen „4 Sterne“ und „5 Sterne“!

### 1.4.3 Metrische Skala

- Numerisch mit interpretierbaren Abständen
- Erlaubte Operationen: Alle arithmetischen Operationen
- Beispiele: Zeit, Gewicht, Umsatz, Temperatur
- Auswertungen: Mittelwert, Varianz/Stdabw, Korrelation, Regression

#### Unterscheidung:

- **Intervallskala:** Kein natürlicher Nullpunkt (z. B. Temperatur in °C, Datum)
- **Verhältnisskala:** Natürlicher Nullpunkt (z. B. Zeit, Gewicht, Geld)

#### Informatik-Beispiele:

Merkmal	Skalentyp	Begründung
Response-Zeit (ms)	Verhältnis	0 ms = keine Zeit, 200 ms = 2× so lang wie 100 ms
Speicherverbrauch (GB)	Verhältnis	0 GB = kein Verbrauch
CPU-Auslastung (%)	Verhältnis	0% = keine Last
Dateigröße (Bytes)	Verhältnis	0 Bytes = leer
Lines of Code	Verhältnis	0 LOC = keine Zeilen
Timestamp (Unix-Zeit)	Intervall	Abstände interpretierbar, Nullpunkt willkürlich
Temperatur (°C)	Intervall	0°C „keine Temperatur“



#### Merksatz:

Nicht jede Zahl ist metrisch (z. B. Postleitzahlen).

### 1.4.4 Typische Fallen in der Informatik

Merkmal	Sieht aus wie...	Ist aber...	Warum?
HTTP-Statuscode (200, 404, 500)	Metrisch	Nominal	404 ist nicht „doppelt so viel“ wie 200
IP-Adresse	Metrisch	Nominal	Keine sinnvolle Ordnung oder Arithmetik
User-ID	Metrisch	Nominal	Rein technischer Identifier
Versionsnummer (2.1.3)	Metrisch	Ordinal	v2.0 < v2.1, aber Abstände nicht definiert
Timestamp	Metrisch	Intervall	Kein absoluter Nullpunkt

### 1.4.5 Übersicht: Erlaubte Operationen

Messniveau	=	< >	+ −	× ÷	Lagemaß
Nominal					Modus
Ordinal					Modus, Median
Intervall					Modus, Median, Mittelwert
Verhältnis					Modus, Median, Mittelwert

## 1.5 Deskriptive vs. induktive Statistik

### 1.5.1 Deskriptive Statistik

Beschreibt den **vorliegenden Datensatz**:

- Häufigkeitstabellen und Diagramme
- Lage- und Streuungsmaße
- Leitfrage: *Wie sehen die Daten aus?*

**Informatik-Beispiele:**

- „Der Median der Response-Zeit war heute 45 ms.“
- „95% der Requests waren unter 200 ms.“
- „Der häufigste Fehler war HTTP 404 (35% aller Errors).“
- „Die CPU-Last schwankte zwischen 20% und 85%.“

### 1.5.2 Induktive Statistik (schließende Statistik)

Schließt von der **Stichprobe auf die Grundgesamtheit**:

- Punktschätzungen
- Konfidenzintervalle
- Hypothesentests

**Informatik-Beispiele:**

- „Mit 95% Konfidenz liegt die wahre mittlere Response-Zeit zwischen 42 ms und 48 ms.“
- „Die neue Version ist signifikant schneller ( $p < 0.05$ ).“
- „Die Conversion-Rate von Variante B ist höher als die der Kontrollgruppe.“
- „Wir können die Nullhypothese ‘kein Unterschied’ ablehnen.“



**Kernidee:** Induktive Aussagen sind nie absolut sicher – sie besitzen ein **Konfidenz- oder Signifikanzniveau**.

### 1.5.3 Vergleich: Deskriptiv vs. Induktiv

Aspekt	Deskriptiv	Induktiv
<b>Fokus</b>	Vorliegende Stichprobe	Grundgesamtheit
<b>Frage</b>	Wie sehen die Daten aus?	Was gilt allgemein?

Aspekt	Deskriptiv	Induktiv
<b>Aussagetyt</b>	Beschreibend	Schließend
<b>Unsicherheit</b>	Keine (Fakten über Daten)	Quantifiziert (KI, p-Wert)
<b>IT-Beispiel</b>	„Median war 45 ms“	„95%-KI: [42, 48] ms“

#### 1.5.4 Praxisbeispiel: A/B-Test

**Szenario:** Ein Online-Shop testet zwei Checkout-Designs.

##### Deskriptive Aussagen:

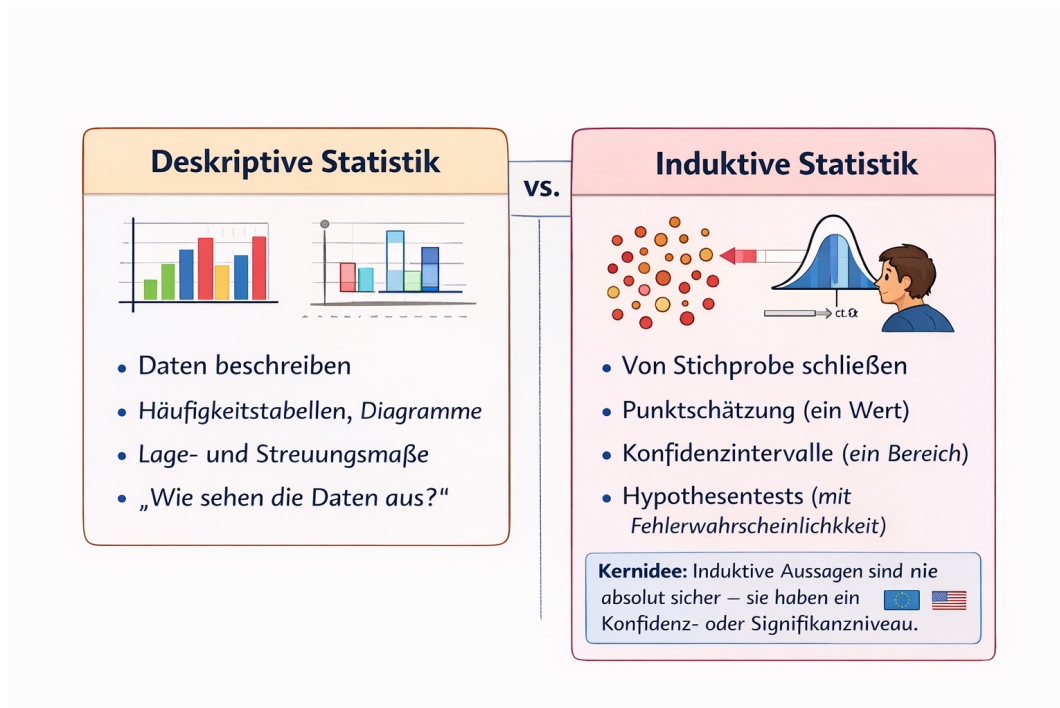
- Variante A: 120 von 1000 Besuchern kauften (12,0%)
- Variante B: 145 von 1000 Besuchern kauften (14,5%)
- Differenz in der Stichprobe: 2,5 Prozentpunkte

##### Induktive Frage:

Ist der Unterschied **statistisch signifikant** oder nur Zufall?

##### Induktive Aussage:

- p-Wert = 0,032 ( $< 0,05$ )
- „Der Unterschied ist statistisch signifikant. Mit 95% Sicherheit ist Variante B besser als Variante A.“



## 1.6 Typischer Statistik-Workflow

1. Fragestellung präzisieren

2. Grundgesamtheit definieren
3. Stichprobenplan / Datenerhebung
4. Daten bereinigen (Fehler, Missing Values, Ausreißer)
5. Deskriptive Auswertung
6. Modellannahmen formulieren
7. Induktiv schließen
8. Interpretation im Kontext

### 1.6.1 Workflow am IT-Beispiel: SLA-Prüfung

**Fragestellung:** Erfüllt unsere API das SLA „95% der Requests unter 200 ms“?

Schritt	Aktivität	Ergebnis
1. Fragestellung	SLA-Compliance prüfen	„95% unter 200 ms?“
2. Grundgesamtheit	Alle Requests des Monats	~30 Mio. Requests
3. Stichprobe	Random Sampling aus Logs	n = 10.000 Requests
4. Bereinigung	Negative Werte, NaN entfernen	n = 9.847 valide
5. Deskription	p95 = 187 ms, 96,2% < 200 ms	Stichproben-Statistik
6. Modellannahme	Log-Normalverteilung	Typisch für Latenz
7. Induktion	95%-KI für Anteil: [95,8%, 96,6%]	Schätzung für Population
8. Interpretation	Untere KI-Grenze > 95%	<b>SLA wird erfüllt</b>

### 1.6.2 Typische Fehler im Workflow

Schritt	Häufiger Fehler	Konsequenz
1	Vage Fragestellung	Keine klaren Schlussfolgerungen
2	Grundgesamtheit unklar	Ergebnisse nicht übertragbar
3	Verzernte Stichprobe	Bias in allen Ergebnissen
4	Keine Bereinigung	Ausreißer verfälschen Statistiken
5	Nur Mittelwert betrachtet	Wichtige Muster übersehen
6	Falsche Modellannahmen	Ungültige Konfidenzintervalle
7	p-Wert falsch interpretiert	Falsche Entscheidungen
8	Kontext ignoriert	Statistisch signifikant    praktisch relevant

## 1.7 Anwendungsbeispiele aus der Informatik

### 1.7.1 Beispiel 1: Performance-Monitoring

**Szenario:** Ein DevOps-Team überwacht Server-Response-Zeiten.

### Grundbegriffe angewendet:

- **Einheit:** Ein einzelner HTTP-Request
- **Merkmal:** Response-Zeit
- **Grundgesamtheit:** Alle Requests der letzten 24 Stunden
- **Stichprobe:** 1% zufällig geloggte Requests
- **Messniveau:** Metrisch (Verhältnis)

### Deskriptive Analyse:

- Mittelwert: 127 ms
- Median: 89 ms
- p95: 342 ms
- p99: 892 ms

### Interpretation:

Der große Unterschied zwischen Mittelwert (127 ms) und Median (89 ms) deutet auf eine rechtsschiefe Verteilung hin – typisch für Response-Zeiten. Wenige langsame Requests ziehen den Mittelwert nach oben.

### 1.7.2 Beispiel 2: A/B-Test für UI-Änderung

**Szenario:** Ein Produktteam testet einen neuen „Kaufen“-Button.

#### Grundbegriffe angewendet:

- **Einheit:** Ein Website-Besucher
- **Merkmal:** Kaufabschluss (ja/nein)
- **Grundgesamtheit:** Alle zukünftigen Besucher
- **Stichprobe:** 10.000 Besucher im Test
- **Messniveau:** Nominal (binär)

#### Ergebnisse:

Variante	Besucher	Käufe	Conversion
A (alt)	5.000	250	5,0%
B (neu)	5.000	290	5,8%

**Induktive Frage:** Ist der Unterschied (0,8 Prozentpunkte) signifikant?

### 1.7.3 Beispiel 3: Log-Analyse für Anomalie-Erkennung

**Szenario:** Ein Security-Team analysiert Login-Versuche.

#### Grundbegriffe angewendet:

- **Einheit:** Ein Login-Versuch
- **Merkmale:**
  - Zeitstempel (metrisch/Intervall)
  - Erfolg (nominal: ja/nein)
  - IP-Adresse (nominal)
  - User-Agent (nominal)
- **Grundgesamtheit:** Alle Login-Versuche
- **Stichprobe:** Echtzeit-Stream der letzten Stunde

#### Statistische Frage:

Wie viele fehlgeschlagene Logins pro Minute sind „normal“? Ab wann ist die Rate so hoch, dass es sich um einen Angriff handeln könnte?

→ Z-Score-Analyse: Werte mit  $|z| > 3$  als Anomalie markieren.

---

## 1.8 Zusammenfassung

- Statistik als **Daten**, **Methode** und **Wissenschaft**
- Zentrale Begriffe: Einheit, Merkmal, Grundgesamtheit, Stichprobe
- Messniveau bestimmt zulässige Auswertungen
- Deskription beschreibt – Induktion schließt (mit Unsicherheit)

### 1.8.1 Zusammenfassung für Informatiker

Konzept	Definition	IT-Beispiel
<b>Einheit</b>	Untersuchungsobjekt	Ein API-Request
<b>Merkmal</b>	Eigenschaft der Einheit	Response-Zeit
<b>Ausprägung</b>	Konkreter Wert	142 ms
<b>Grundgesamtheit</b>	Alle relevanten Einheiten	Alle Requests eines Monats
<b>Stichprobe</b>	Beobachtete Teilmenge	10.000 geloggte Requests
<b>Deskriptiv</b>	Beschreibt Daten	„Median war 45 ms”
<b>Induktiv</b>	Schließt auf Grundgesamtheit	„95%-KI: [42, 48] ms”

### 1.8.2 Messniveau-Cheatsheet für IT-Merkmale

Messniveau	Erlaubte Maße	IT-Beispiele
<b>Nominal</b>	Modus, Häufigkeiten	HTTP-Methode, Browser, OS
<b>Ordinal</b>	+ Median, Quartile	Bug-Priorität, Log-Level
<b>Metrisch</b>	+ Mittelwert, Varianz	Response-Zeit, CPU-Last

### 1.8.3 Wichtige Prinzipien

1. **Repräsentativität:** Stichprobe muss die Grundgesamtheit abbilden.
  2. **Messniveau beachten:** Nicht jede Kennzahl ist für jedes Merkmal sinnvoll.
  3. **Unsicherheit quantifizieren:** Induktive Aussagen haben immer eine Fehlerwahrscheinlichkeit.
  4. **Mehrere Kennzahlen:** Nie nur den Mittelwert betrachten!
  5. **Kontext beachten:** Statistisch signifikant    praktisch relevant.
- 
- 

## 1.9 Linksammlung

### Notebooks

[https://colab.research.google.com/github/karkessler/dhbw-statistik/blob/main/notebooks/kapitel\\_01\\_einfuehrung.ipynb](https://colab.research.google.com/github/karkessler/dhbw-statistik/blob/main/notebooks/kapitel_01_einfuehrung.ipynb)

### Übungen

<https://github.com/karkessler/dhbw-statistik/blob/main/pdf/uebungen/uebungsbblatt1.pdf>

---