

# 1. Übung: Datenerhebung & Messung

Yichen Han



26. Oktober 2023

- 1 Organisation
- 2 Wiederholung
- 3 Blatt 1

- ① **Vor der Übung die Aufgaben selbstständig versuchen zu lösen.**
- ② Abgabe der Hausarbeiten (Ende jedes Monats) in Gruppen erwünscht.
- ③ Es wird **keine** Musterlösung inkl. Folien hochgeladen.

- ① **Vor der Übung die Aufgaben selbstständig versuchen zu lösen.**
- ② Abgabe der Hausarbeiten (Ende jedes Monats) in Gruppen erwünscht.
- ③ Es wird **keine** Musterlösung inkl. Folien hochgeladen.
- ④ Interaktive Lehre mithilfe der Plattform Particify (kostenlos, keine Registrierung erforderlich, anonym)



- ⑤ Fragen & Feedbacks gerne per Email an [Yichen.Han@campus.lmu.de](mailto:Yichen.Han@campus.lmu.de)  
Gerne bin ich auch Ihre Kontaktperson für die Fachschaft.

- ① kleines Quiz für die Vorlesung
- ② kurze Wiederholung der wesentlichen Inhalte der Vorlesung
- ③ Darstellung der Lösungen zu Übungsaufgaben
- ④ ggf. kleine interaktive Aufgaben oder Zeit für Fragen
- ⑤ wenn Zeit knapp ist → Aufgaben und Wiederholung parallel

- ① Die Übungsblätter bis Ende des 2. Semesters aufbewahren.
- ② Mathematische Kenntnisse auffrischen und ggf. erweitern.
- ③ Das kostenlose Konto von DataCamp gut nutzen.
- ④ Das Studium sinnvoll digitalisieren. (GitHub,  $\text{\LaTeX}$ , Stackoverflow, ChatGPT, usw.)

- 1 Organisation
- 2 Wiederholung**
- 3 Blatt 1



Quiz 1:





Quiz 1:

- ➊ Grundbegriffe: Grundgesamtheit, Untersuchungseinheit, Merkmal, Ausprägung, Beobachtung
- ➋ Skalenniveaus: nominal, ordinal, intervall, verhältnis, absolut
- ➌ Merkmalstypen: stetig, diskret, quasi-stetig
- ➍ Erhebungsarten (Methode, Datenform, Umfang): Experiment, Befragung, Beobachtung, Vollerhebung, Stichprobe, Querschnittsdaten, Zeitreihe, Längsschnittdaten

Warum sind Querschnittsstudien nicht in der Lage, eine (zeitliche) Tendenz aus den Daten zu verdeutlichen? (Kohorteneffekt)<sup>1</sup>:

## **Szenario:**

Eine Querschnittsstudie untersucht die Nutzung von TikTok der verschiedenen Altersgruppen im Jahr 2023. Die Studie findet heraus, dass ältere Menschen tendenziell weniger TikTok bevorzugen.

Diese Schlussfolgerung ist aber nicht sinnvoll.

---

<sup>1</sup>Thoughtco, Was ist ein Kohorteneffekt? Definition und Beispiele

- 1 Organisation
- 2 Wiederholung
- 3 Blatt 1

## Die drei Säulen der Statistik

- **Deskriptive:** Beschreibung von betrachteten Merkmalen, graphische Datenaufbereitung, liefert erster Eindruck der Daten, und hilft bei Datenvalidierung (erkennt Fehler). Keine Rückschlüsse auf die Grundgesamtheit über Erhebungsdaten möglich.

## Die drei Säulen der Statistik

- **Deskriptive:** Beschreibung von betrachteten Merkmalen, graphische Datenaufbereitung, liefert erster Eindruck der Daten, und hilft bei Datenvalidierung (erkennt Fehler). Keine Rückschlüsse auf die Grundgesamtheit über Erhebungsdaten möglich.
- **Explorative:** Suche nach Strukturen in den Daten (ohne stochastische Methoden), Formulierung von Hypothesen für das den Daten zugrunde liegende stochastische Modell (wichtig für die Induktion).

## Die drei Säulen der Statistik

- **Deskriptive:** Beschreibung von betrachteten Merkmalen, graphische Datenaufbereitung, liefert erster Eindruck der Daten, und hilft bei Datenvalidierung (erkennt Fehler). Keine Rückschlüsse auf die Grundgesamtheit über Erhebungsdaten möglich.
- **Explorative:** Suche nach Strukturen in den Daten (ohne stochastische Methoden), Formulierung von Hypothesen für das den Daten zugrunde liegende stochastische Modell (wichtig für die Induktion).
- **Induktive:** Ziehung von Schlüssen von der Stichprobe auf die Grundgesamtheit; Verwendung entsprechender statistischer Methoden auf Basis eines wahrscheinlichkeitstheoretischen Modells.  
(Punktschätzung, Konfidenzintervalle, Tests)

Bestimmen Sie die Grundgesamtheit ( $\Omega$ ), die statistischen Einheiten ( $\omega_i$ ) und Beobachtungen, die untersuchten Merkmale ( $X$ ), die theoretisch möglichen Merkmalsausprägungen ( $S$ ), und weitere erfassbare Merkmale.

- ①  $\Omega$  : Residenzkonzerte in München im Jahr 2015;  
 $\omega_i$  : Residenzkonzert  $i$  in München im Jahr 2015;  
 $X$  : Anzahl der verkauften Karten bei einem Residenzkonzert;  
 $S = \mathbb{N}_0$ ;  
weitere: Art und Datum des Konzertes, usw.
- ②  $\Omega$  : Studierende der LMU;  
 $\omega_i$  : Studierende(r)  $i$  an der LMU;  
 $X$  : primäres Verkehrsmittel, das zur Fahrt zur Uni genutzt wird;  
 $S$  : Menge aller möglichen Verkehrsmittel;  
weitere: Studienfach, Alter, Geschlecht der Studierenden usw.

- ③  $\Omega$  : Amtlich betriebene Stationen zur Messung von Luftschadstoffen in Bayern;  
 $\omega_i$  : Amtlich betriebene Station  $i$ ;  
 $X = (x, y)$ : GPS-Koordinate einer Station;  
 $S = [47, 51] \times [9, 14]$ ; <sup>2</sup>  
weitere: Baujahr der Station, Ballungsraum ja/nein, Entfernung zur nächsten Industrieanlage, usw.

---

<sup>2</sup>GPS-Bereich von Bayern, müssen Sie nicht wissen





Quiz 2:

	Natürliche Reihen- folge	Interpretierbare Differenzen	Natürlicher Null- punkt	Natürliche Einheit
Nominalskala	nein	nein	nein	nein
Ordinalskala	ja	nein	nein	nein
Intervallskala	ja	ja	nein	nein
Verhältnisskala	ja	ja	ja	nein
Absolutskala	ja	ja	ja	ja

**Tabelle:** Charakterisierung der Skalenniveaus

Erläutern Sie geeignete Erhebungsarten (Methode, Datenform, Umfang) für die folgenden Sachverhalte.

- ① Testen eines Düngemittels: *Experiment, Längsschnittanalyse, Stichprobe*
- ② Einschätzung der Fahrtüchtigkeit: *Befragung, Querschnittanalyse, Stichprobe*
- ③ Schätzung der durchschnittlichen Lebensdauer von Leuchtstoffröhren: *Beobachtung, Längsschnittanalyse, Stichprobe*

Welches Auswahlverfahren würden Sie wählen?

## Szenario 1:

Eine Untersuchung soll Aufschluss über den **durchschnittlichen Quadratmeterpreis** von Mietwohnungen in einer Stadt geben. Sie wissen, dass man die Stadt in **drei „Regionen“**; einteilen kann, in denen die Quadratmeterpreise jeweils ähnlich sind.

Welches Auswahlverfahren würden Sie wählen?

## Szenario 1:

Eine Untersuchung soll Aufschluss über den **durchschnittlichen Quadratmeterpreis** von Mietwohnungen in einer Stadt geben. Sie wissen, dass man die Stadt in **drei „Regionen“**; einteilen kann, in denen die Quadratmeterpreise jeweils ähnlich sind.

- einfache Zufallsstichprobe? Mietpreise stark von der „Region“ abhängig.

Welches Auswahlverfahren würden Sie wählen?

## Szenario 1:

Eine Untersuchung soll Aufschluss über den **durchschnittlichen Quadratmeterpreis** von Mietwohnungen in einer Stadt geben. Sie wissen, dass man die Stadt in **drei „Regionen“**; einteilen kann, in denen die Quadratmeterpreise jeweils ähnlich sind.

- einfache Zufallsstichprobe? Mietpreise stark von der „Region“ abhängig.
- -> jeweils eine Stichprobe mit dem gleichen Umfang in jeder Region durchführen.

Vorteile: Wissen über jede Region + niedrigere Varianz für das Gesamtmittel (bessere Schätzung)

## Geschichtete Stichprobe

Referenz: Fahrmeir et al. Kap.1.4 (S. 23-24)

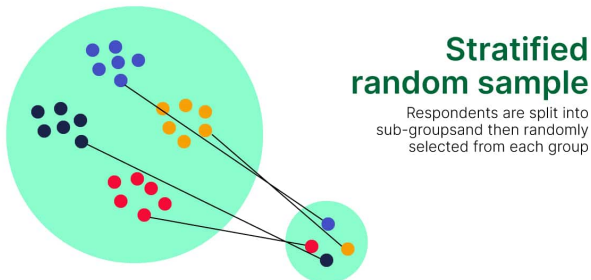


Abbildung: Beispiel der geschichteten Stichprobe

Effizienzgewinn, falls die relevanten Merkmale innerhalb der Schichten homogen und zwischen den Schichten heterogen sind.

## Szenario 2:

Zur Bestimmung des Zigarettenkonsums von Hauptschülern in der 8. Klasse soll eine Erhebung mit Hilfe von Fragebögen durchgeführt werden.

- Ist es möglich, in einem Jahrgang eine Stichprobe durchzuführen?
- Ist es möglich, eine Vollerhebung durchzuführen?
- Was muss zufällig ausgewählt werden?

## (Einstufige) Klumpenstichprobe

Referenz: Fahrmeir et al. Kap.1.4 (S. 23-24)

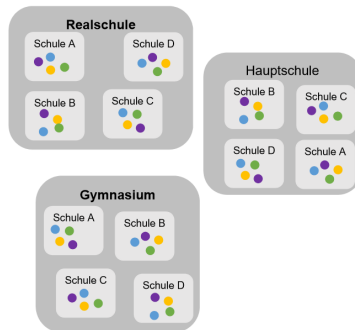


Abbildung: Beispiel der Klumpenstichprobe, Source

- -> In ein paar zufällig ausgewählten Hauptschulen jeweils eine Vollerhebung für Schüler:innen der 8. Klasse durchführen.



## Grundlagen der Wahrscheinlichkeitsrechnung Blatt 2 für Abgabe:

- **Ausschließlich** die erste und zweite Aufgabe werden benotet!
- Die Gruppe erhält die gleiche Bewertung.
- Achten Sie auf die Fristen!