# Statistische Methoden in der Geographie + VL 1

23.10.17
Klaus Sachs

## Definition Statistik:

„Eine wissenschaftlich betriebene Statistik befähigt uns, Beobachtungen hinlänglich zu beschreiben und allgemeine Schlussfolgerungen zu ziehen. Sie gestattet überdies, den Geltungsbereich und Verallgemeinerungswürdigkeit empirischer Befunde zu beurteilen."
→ Die Statistik bemüht sich Gesetze bzw. Gesetzmäßigkeiten zu erkennen, deren eine Verteilung von Merkmalen folgt

## Statistische Erkenntnisse auch im Alltag:

- Verallgemeinerung von Einzelerfahrung führt zu (statistischem) Wissen und beeinflusst (unser) Handeln. (Allerdings besteht die Gefahr sich zu irren)
- Bei wissenschaftlich betriebener Statistik: Intuition und lernen durch Erfahrung / Versuch wird durch mathemathische Prinzipien und Rechenverfahren ersetzt

## Man unterscheidet 3 Teilbereiche:

### 1. Deskriptive Statistik:

1 Variable

- Dient der Beschreibung der Verteilung von Merkmalen
- z. B durch Bilden von Durchschnittswerten, bestimmten Häufigkeiten oder Aussagen über die Streuung eines Merkmals
- Durch Ordnen, Gruppieren + Verdichten werden große Datenmengen übersichtlicher // absolute, relative Häufigkeiten // Lage- und Streuungsmaße (1 Beschreibung)
- Erleichtert das Erkennen vom Charakteristischen, Wichtigen in einer Verteilung (Tabellen, Graphiken, Klassenbildung durch Wählen von Bereichen, Angaben von Streuungen)
→ Statistiken fassen zusammen

### 2. Schätzstatistik

2 Prognose auf Grund einer Schätzung

- Gibt an mit welcher Wahrscheinlichkeit bestimmte Parameter, die man für eine Stichprobe berechnet hat, auch die Parameter der Grundgesamtheit wiederspiegeln // Erlaubt auf Grund von wenigen Daten einer Teilmenge (Stichprobe) Voraussagen (Schätzungen, Prognosen) für die Gesamtmenge (Grundgesamtheit)
→ Wahlprognosen

### 3. Prüf- und Teststatistik

3 Welcher Zusammenhang besteht zw. 2 Variablen → 2 Variablen

- (auch operative, schließende Statistik)
- Hilft Zusammenhänge zwischen verschiedenen (2 oder mehr) Merkmalen aufzudecken
- kam in Planungsprozessen zur Entscheidungsfindung genutzt werden
  z. B. Zusammenhang zw. Mobilität und Verkehrsaufkommen (.. und Umwelt.)

# Statistische Methoden in der Geographie – VL 2

30.10.17

## Definition Statistik

→ Die Statistik ist ein wissenschaftliches Werkzeug, dass sich der Mensch geschaffen hat, um von der verwirrenden Vielfalt beobachtbarer Erscheinungen zu deren Wesen vorzudringen und die Gesetze der objektiven Realität dadurch tiefer erkennen zu können.

→ Auf diese Weise befähigt uns eine wissenschaftlich betriebene Statistik „Beobachtungen hinlänglich genau zu beschreiben und allgemeine Schlussfolgerungen zu ziehen. Sie gestattet überdies den Geltungsbereich und die Verallgemeinerungswürdigkeit empirischer Funde zu beurteilen."

Exakte Beschreibung (Darstellung) empirischer Daten
↓
Erkennen von Strukturen + Gesetzmäßigkeiten
↓
Schlussfolgerungen, Verallgemeinerungen (Prognosen)

## Stellenwert statistischer Verfahren

- in der Öffentlichkeit
  exakte Methode
  hoher 'Wahrheitsgehalt'
  aber auch Lügen mit Statistik

- in der Wissenschaft
  wissenschaftliches Instrument, um beobachtbare Erscheinungen mit Hilfe mathematischer Prinzipien und Rechenverfahren zu strukturieren und zu erklären

## Wissenschaftstheorie

### Was ist die Aufgabe von Wissenschaft?

- bemüht sich darum vielfältige Ereignisse in Natur / menschlichem Zusammenleben zu sammeln, ordnen + Aussagen über innere Verbundenheit zu machen
- aus analytisch-szientistischer Sicht:
  → Wissenschaft versucht mit Hilfe von methodisch betriebener Forschung allgemeine Gesetze + Theorien aufzubauen → tragen dazu bei reale Welt zu erkennen + zu erklären

## Anspruch an Wissenschaft

- Resultate von Forschungsaktivitäten sollen nicht nur von denen die sie erbringen als richtig anerkannt werden, sondern für alle Beteiligten + Interessierten akzeptierbar sein (wahr) → mit der Realität übereinstimmen.

**ABER:** Was ist wahr? Realität? Inwieweit können wir mit unseren bestechlichen Sinnesorganen Aussagen über eine objektive Realität machen

Sie leben in einer Höhle durchs angebunden sein können sie sich nicht informieren Projektionen an Wand = Wahrheit werden sie befreit möchten sie wieder zurück in ihre Höhle Außenwelt ≠ Wahrheit

Bsp: Die reale Welt (Wahrheit), ist die Welt wie sie uns erscheint (Platos Höhlengleichnis)

**Fazit:** Sinnesorgane nehmen die Umgebung unterschiedlich wahr (je nach Stimmung)
... Farbwahrnehmung von Individuen, Insekten, Rot-Grün-Blinden
... Geräusche einer Straßenbahn nach Techno-Nacht angenehm leise, am Sommerabend laut!

Individuelle Wahrnehmungen hängen zudem vom Vorwissen, Erfahrungen und Erwartungen ab
... Japaner sehen HD anders als HDler
... Erstis sehen Alpen anders als 8. Semester

**Empirismus und Positivismus (Beobachtungen) führen nicht zwangsläufig zu wahren Aussagen.**

30.10.17 II

## Welche Rolle spielt 'Wahrheit' in der Wissenschaft?

- auch in Wissenschaft nur Vermutungen, Hypothesen → keine absoluten Wahrheiten
- Hypothesen + Theorien sind empirisch mehr oder weniger gut bestätigte Aussagen über die Realität
- Bleiben immer vorläufige, relative Wahrheiten
- Anspruch einer absoluten Wahrheit = unwissenschaftlich

⟶ Aber: alle Wissenschaften haben die Aufgabe nach der Wahrheit zu suchen, nach Theorien, die der Kritik standhalten

## Wie kann sich Wissenschaft dem Phänomen 'Wahrheit' nähern?

- Grundprinzip von Wissenschaft: bedingungslose Kritik → prinzipielle Zweifel an jeder Aussage und Behauptung
- Prinzip wurde im 17. Jhd von franz. Philosoph Rene Descartes in die Wissenschaft eingeführt
- Eine wissenschaftliche Revolution (stellt die Prinzipien der Kritik + des logischen Denkens + des empirischen Prüfens in den Mittelpunkt)
  → Alles als Behauptung, als Hypothese, als vorläufige, relative Wahrheit

↳

## Kritischer Rationalismus und Prinzip der Falsifikation

→ Carl R. Popper: alles Wissen behält stets einen provisorischen, hypothetischen Charakter (auf Grundlage von Platos Erkenntnis (dem Höhlengleichnis)).

→ Blotevogel: Jede Behauptung, jede Hypothese, jede Theorie ist prinzipiell ständig der Kritik auszusetzen → in Frage zu stellen (es muss immer die Möglichkeit der Kritik bestehen...

→ Für Popper: Prinzip der Falsifikation = entscheidendes Prüfkriterium
- im Unterschied zum traditionellen Positivismus (sieht Verifikation und Falsifikation als gleichwertige Möglichkeiten der Hypothesenführung ansicht) argumentiert Popper: Die Verifikation ist aus logischen Gründen unzureichend sie führt zur unkritischen Bestätigung von Hypothesen.

→ Wie funktioniert das Prinzip der Falsifikation?
- Hypothesen werden nicht durch Verifikation bewiesen – ABER: durch Falsifikation belegt werden: Beobachtungen werden aufgewiesen, die zur Folgerung im Widerspruch stehen, die aus der Hypothese abzuleiten sind.
- ist entscheidend so exakt wie möglich zu arbeiten und die Theorien (unsere Erfindungen / unsere Konstruktionen) möglichst strengen Prüfungen zu unterziehen
- Möglichkeit, solche strengen Prüfungen von Theorien (relativ exakt) vorzunehmen, ist d. Einsatz statistischer Methoden. Basieren auf der Verwendung mathematischer Prinzipien (Prüf- und Teststatistik)

## Konsequenzen für sie als Wissenschaftler/innen

Sie sollten...

... ihr eigenes tun transparent machen, damit es andere nachvollziehen können und sich ein Bild über die Einschränkungen und Grenzen des Aussagen machen zu können.

... ihr eigenes tun ständig der Selbstkritik auszusetzen (reflexive Ebene)

BRUNNEN

Statistische Methoden in der Geographie - VL 3 06.11.17

## Quantitative und Qualitative Methoden

## Empirische Methoden in Ethnologie und Geographie

beobachtungsorientiert

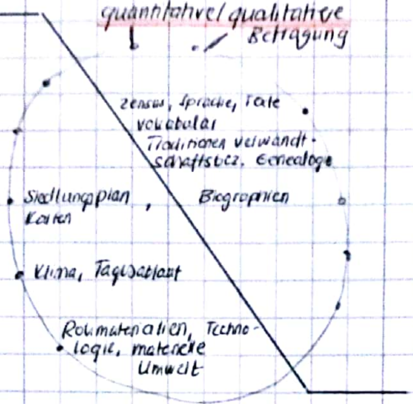
geographische Beobachtungen

teilnehmende Beobachtungen

Kartierung

Luft- und Satellitenbilder

Auswertung von Stadtplänen




informationsorientiert

Auswertung von Primär und Sekundär dateien

Auswertung von Quellen und Dokumenten

Ermittlung von Biographien

Tests

Labormethoden

### Quantitative Verfahren

- Datenerhebung standardisiert
- Antwortkategorien vorgegeben (inhaltliche Vergröberung)
- überschaubare, in standardisierten Kategorien geordnete Datenmenge
- hohe Repräsentativität durch umfangreiches Sample (große Zahl an Fragebögen, Beobachtungen...)
- intersubjektive Überprüfbarkeit möglich

Konsequenz daraus:

SCHEMATISIERUNG

- geeignet für die Erhebung 'harter' Daten und kategorisierbarer Informationen
- Auswertung mit objektiven, normierten, mathematisch-statistischen Verfahren
- Dokumentation der Daten ist unproblematisch

### Qualitative Verfahren

- Datenerhebung nicht (kaum) standardisiert
- Antwortkategorien nicht vorgegeben (ausführliche, detaillierte, nuancenreiche Antworten)
- nicht (kaum) wenig strukturierte Datenfülle
- keine Repräsentativität (im statistischen Sinne), da nur wenige Einzelfälle erfasst
- intersubjektive Überprüfbarkeit nicht m.

INDIVIDUALISIERUNG

- geeignet für eine differenzierte Untersuchung von Einzelfällen, Prozessanalysen etc. erhoben werden Meinungen, Einstellungen, Images...
- Auswertung mit interpretativen Verfahren (subjektive, nicht normierte Einflüsse)
- Dokumentation der Daten ist schwierig, problematisch z.T. nicht möglich

### Semantisches Differential

aus Gabler Wirtschaftslexikon

Skalierungsverfahren zur Messung des Images von Objekten, Personen etc. (Einstellung). Die Versuchspersonen versuchen stufen auf Bewertungsskalen (Rating Skalen, Rating) ein Untersuchungsobjekt ein. Die beiden Pole jeder Skala stellen verbale Gegensätze dar. Die Abstufungen bleiben verbal undefiniert und weisen optisch gleiche Abstände auf. Zur Auswertung des semantischen Differentials dient neben der Mittelwertbildung und Streuungsberechnung über die Menge der Testpersonen die Methode der Datenreduktion; graphische veranschaulichung durch Darstellung der jeweiligen Profile.

VORSICHT: Durch Befragungsstandort kann es zu einer Verzerrung der Ergebnisse kommen. Menschen, die am Bahnhof statt Flughafen befragt wurden, sagen eher, dass HD teuer ist. Viele junge Leute werden befragt = HD = eher teuer.

BRUNNEN

### Foto - Erkennungs / Assoziations - Test

### Mental maps (kognitive Karten)

## Primär- und Sekundärstatistik

| Primärstatistische Daten | Sekundärstatistische Daten |
|---|---|
| - Datenmaterial wurde eigens für den jeweiligen Zweck erhoben (d.h. eigene Erhebung)<br>- spezielle Befragung zu bestimmten Themen. Vorteile: Flexibilität in der Gestaltung der Datenbasis, Zuverlässigkeit und Aktualität der Daten.<br>- Nachteile: Zeit und kostenintensiv<br>Bsp.: Fragebögen für eigene Untersuchung<br>→ (ALLBUS-Daten, vollständige Datensätze erhältlich beim Zentralarchiv für empirische Sozialforschung, Köln) | - Datenmaterial wurde ursprünglich für anderen Zweck erhoben, als für den es eingesetzt wird (aufbereitete Primärdaten)<br>- bestehende Daten werden zur Untersuchung herangezogen.<br>- Vorteil: geringe Kosten und geringer zeitlicher Aufwand, durch Verknüpfung v. bestehenden Daten eine Verbreitung der Datenbasis<br>- Nachteil: nicht gezwungen aktuell<br>Bsp.: Volkszählungsdaten, Mikrozensus (0,1 oder 1% der Bevölkerung werden befragt), in Auszügen erhältlich bei den jeweiligen Landesämtern<br>Mikrozensus 830 000 Personen, 370.000 Haushalte<br>(Statistische Jahrbücher (national und international), nationale und internationale Behörden z.B. BMZ, EU, FAO, UNO etc.<br>→ sobald es sich um aufbereitete Daten in Tabellen und nicht mehr um Einzelfälle (Einzelne Fragebögen) handelt)<br>→ Sekundärstatistik |

Zuvor → vollständige Datensätze man weiß nicht die exakte Straße z.B. aber es geht noch um Einzelfälle

## Datengrundlage

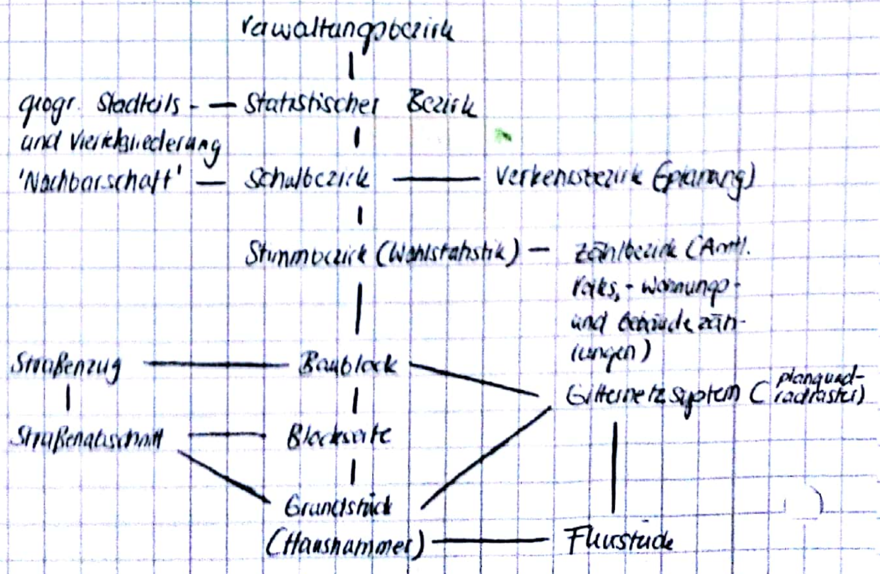
| Eigenständige Erhebung (Primärstatistische Daten) | Nutzung vorhandener Daten (Sekundärstatistische Daten) |
|---|---|
| - Befragung (Haushalte, Passanten, Besucher)<br>- Beobachtungen<br>- Kartierung (Brachflächen, landwirtschaftliche Nutzflächen)<br>- Messung (Niederschlag, Einstellungen)<br>- Zählungen (Passantenströme) | - amtliche Statistiken (Statistisches Bundesamt, statistische Landesämter)<br>- sonstige Statistiken (Meinungsforschungsinstitute)<br>- Luft- und Satellitenbilder<br>- thematische Karten<br>- sonstige Quellen (Archive, wissenschaftliche Studien) |

## Bezugsquellen

- Sekundärstatistische Daten werden durch stat. Ämter vertrieben (statistische Ämter verschiedener administrativer Raumeinheiten: Bundes- und Landesämter für Statistik, statistische Abteilung von Städten, Gemeinden, Kreisen etc.
- Raumbezogene Daten liefert die laufende Raumbeobachtung des Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR). www.bbsr.bund.de

→ Raumbezogene Daten = besonders wichtig für Geographen

Ab dem Straßenzug gibt es keine Sekundärstatistischen Daten der Landesämter mehr. - auf Grund von Datenschutz dürfen statistische Ämter nicht mehr auswerten



## Darstellungsmöglichkeiten statistischer Informationen

- Text — Balkendiagramm (bar chart) gestapelt
- Tabelle — Säulendiagramm + Liniendiagramm (line chart)
- Tabelle (absolute, relative und kumulierte Werte) (Zahlen F 25 rand → kumul. W)
- Häufigkeiten

### Absolute Häufigkeiten $f_i$

- Anzahl der Elemente mit Merkmalswert $x_i$ (Anzahl der Elemente welche die Merkmalsausprägung [i] besitzen).

- kleinster möglicher Wert = 0 (Merkmalsausprägung kommt in der Stichprobe nicht vor)
- größt möglicher Wert = N (alle Elemente haben die gleiche Merkmalsausprägung)

$0 \leq f_i \leq N$

### Relative Häufigkeiten $h_i$

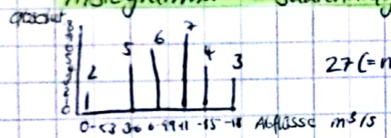
- $h_i = f_i / N$, Dividieren der absoluten Häufigkeit $h_i$ durch die Gesamtanzahl der Elemente (N). → Bei Multiplikation mit 100 erhält man den prozentualen Anteil einer Merkmalsausprägung einer Stichprobe

- kleinster möglicher Wert = 0 (0%)
  " – " –
- größt möglicher Wert = 1 bzw 100%
  " – " –

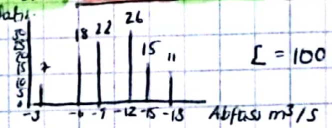
$0 \leq h_i \leq 1$

- Diagrammdarstellung

Histogramm = Säulendiagramm der absoluten oder relativen Häufigkeiten




27 (= n)



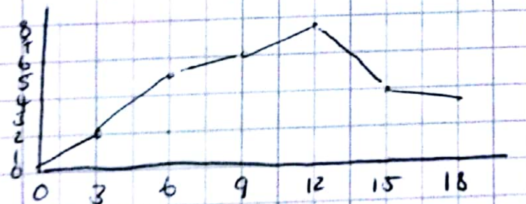

$\Sigma = 100\%$

### Häufigkeits- und Summenpolygon (absolut)

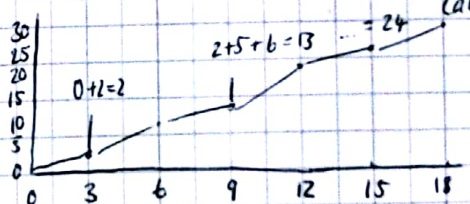
Abflussmengen (27 Messungen)

3 m³/s – 2 7% 12 m³/s – 7 26%
6 m³/s – 5 18% 15 m³/s – 4 15%
9 m³/s – 6 22% 18 m³/s – 3 (Fälle) 11%

Häufigkeitspolygon (absolut)




Polygon der Summenhäufigkeit (kummulierte #.) (absolut)




### Summenpolygon (relativ) → Polygon der relativen Summenhäufigkeit

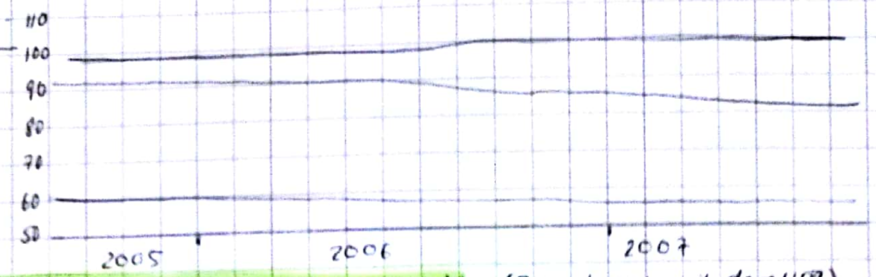

Welcher Anteil der Stichprobe unterschreitet 10 m³/s = ~ 57%

Welcher Anteil überschreitet 14 m³/s = ~ 13%

## Liniendiagramm (mit Index)

Preisindex für Telekommunikationsdienstleistungen 2000 = 100

Index ← (Ausgangsbasis für relative Veränderungen)



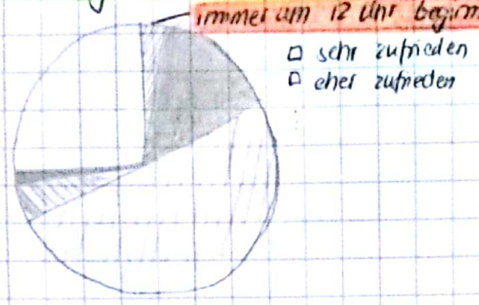

Preisentwicklung stabil

Rückgang um 10%.

um 5%.

## Kreissektorendiagramm (pie chart) (Zufriedenheit mit dem UPR)

immer um 12 Uhr beginnen




Erstellung: $n = 360°$

Dreisatz

$100\,\% = 360°$ } relativ

$1\,\% = 3{,}6°$

$N = 360°$ } absolut

$1\text{ Element} = x$

## Chloropiethenkarte

- flächenhafte Darstellungen flächenbezogener quantitativer Daten
- Einzelflächen mit der Fläche zugeordneten gleichen Werten scharf voneinander abgegrenzt und durch Schraffuren, Farbabstufungen u. a. dargestellt. (ohne Kasten im Bild)
- Beispiel enthält zusätzliche Elemente eines Karto- bzw. Kartodiagramms (gelbe Quadrate)
  → zur Angabe des Flächenumfangs

## Von der Urliste zur Graphik

**1. Urliste**

18, 28, 26, 11, 24, 16, 22, 21, 18,
16, 20, 21, 28, 25, 24, 22, 21, 20,
18, 15, 20, 24, 23, 24, ...

**2. Primäre Tafel**

28, 28, 27 27 27, 26, 26,
26, 26, 26, 25, 25, 25, 25, 24,
24, 24, 24, 24, 24, ...

**3. Häufigkeitsverteilung**

| $x_i$ Punkte | $f_i$ |
|---|---|
| 28 | 2 |
| 27 | 3 |
| 26 | 5 |

$\sum f_i = 72$

**4. Klassenbildung:** Prozess der Klassifizierung bzw. Gruppierung der Daten: Zur Bestimmung der Klassen geht man von einem Intervall aus, das alle Werte umfasst. Dabei sollen sich:

- die verschiedenen Klassen nicht überdecken
- das gesamte Werteintervall von den Klassen überdeckt wird
- die Klassenintervalle, möglichst gleich groß sein / d.h. äquidistant
- die Klassenmitten- und Klassengrenzen sollen möglichst leicht zu merkende Zahlen sein.

↳ **Klassenbreite**: Sind die einzelnen Intervalle in der Häufigkeitsverteilung bei statistischen Erhebungen. Klassenbreite 5 = 1-5, 6-10, ..

## Klassifizierung quantitativ räumlicher Daten (Dr. Christoph Mager)

**Klassenzahl ist abhängig von:**

- Verteilung der Datenwerte
- Zahl der räumlichen Einheiten
- Reproduktionstechniken (Graustufen, Farben)

nach Will (1970) Maximale Klassenanzahl = $\sqrt{\text{Anzahl der Werte}}$

nach Sturgis in Bahrenberg, Giese und Nipper 1990: Klassenzahl $= 1 + 3{,}32 * \log(\text{Anzahl})$

nach Davis 1974: Maximale Klassenzahl $= 5 * \log(\text{Anzahl der Werte})$

## Methoden der Klassenbildung I

Die Klassengrenzen sind frei bestimmbar, in der Regel wird aber auf eine begründbare Methode zurückgegriffen werden, um dem Vorwurf der Manipulation und Willkür zu entgehen.

## Methoden der Klassenbildung II

Jeder Datenwert liegt in einer eigenen Klasse und enthält eine eigne Farbe.

## Methoden der Klassenbildung III: Quantile

Alle Klassen werden gleichmäßig besetzt.

- Vorteile:
  - differenziertes Kartenbild
  - gute Klassifizierung für Entwürfe
- Nachteile:
  - nicht sachlogisch begründbar
  - nur auf eine bestimmte Karte anwendbar
  - Klassenbreiten können sehr unterschiedlich sein

## Methoden der Klassenbildung IV: äquidistante Klassen

Alle Klassen haben die gleiche Breite. Die Klassenzahl ist frei wählbar.

- Vorteile:
  - Unabhängig von der Verteilung der Datenwerte
  - gut geeignet für Vergleiche über Raum und Zeit
- Nachteile:
  - Daten müssen relativ gleichmäßig verteilt sein

## Methoden der Klassenbildung V: Natürliche Brüche

Minimierung der Unterschiede innerhalb einer Klasse, Maximierung der Unterschiede zwischen den Klassen

- Vorteile:
  - gute Representation der Charakteristika der Datenverteilung
  - statistisch begründbar
- Nachteile:
  - nur auf eine bestimmte Karte anwendbar
  - Klassenbreiten können sehr unterschiedlich sein

## Methoden der Klassenbildung VI: Sinnklassen

Klasseneinteilung nach einem Ordnungsprinzip, Klassenzahl kann frei gewählt werden.

- Vorteile:
  - unabhängig von der Verteilung der Daten sachlogisch begründbar
- Nachteile:
  - selten anzuwenden
  - kaum vergleichbare Normen

## Methoden der Klassenbildung VII: arithmetisches Mittel

Obere Klassengrenzen ergeben sich aus dem arithmetischen Mittel der jeweiligen Hälften, Viertel etc.

- Vorteile:
  - Orientierung an der Datenverteilung
  - gut für Vergleiche verschiedener Themen
- Nachteile
  - es können nur 2, 4, 8 etc. Klassen gebildet werden

## Methoden der Klassenbildung VIII: logarithmische Klassen

Klassenbreite nimmt exponentiell zu oder ab

- Vorteile
  - geeignet für Daten mit großer Spannweite oder ungleichmäßiger Verteilung
- Nachteile:
  - selten anzuwenden

BRUNNEN

## Statistische Methoden in der Geographie VL 4 13.11.17

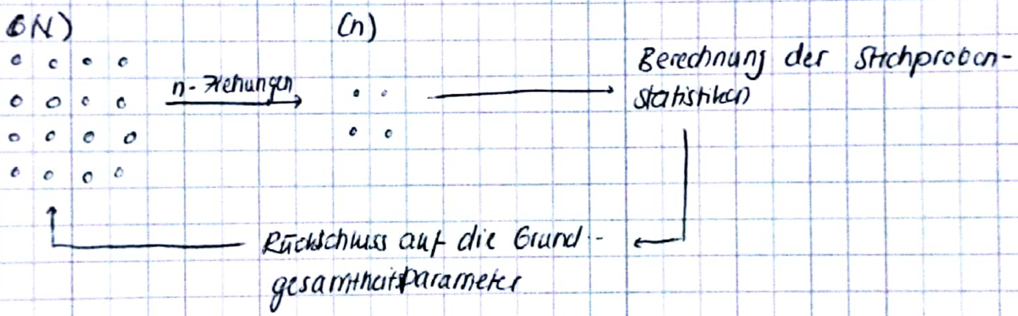
**Die Stichprobe:** (Repräsentativität, intuitives Konzept)

- Auswahl einer Teilgesamtheit ist so zu wählen, dass aus dem Ergebnis der Teilerhebung möglichst genau und sicher auf die Verhältnisse der Gesamtmasse geschlossen werden kann.
- gegeben wenn die Teilerhebung in der Verteilung interessanter Merkmale gemäß der Gesamtmasse ist → verkleinertes aber wirklichkeitsgetreues Bild der Gesamtmasse   Bereboren
- Repräsentativität einer Teilgesamtheit liegt vor, wenn sich in bestimmten Merkmalen eine ähnliche Struktur aufweist wie die Grundgesamtheit. Nur dann kann man von Teil- auf Grundgesamtheit schließen.

**Grundgesamtheit (N):**
- Menge aller Elemente, über die wir durch Untersuchung Erkenntnis gewinnen wollen. Die nicht immer bekannte Anzahl dieser Elemente is[N]

**Stichprobe (n):**
- Stichprobenergebnis informiert über die zur Stichprobe gehörenden Elemente und ~~gestattet~~ mehr oder minder zuverlässige Aussagen über die Grundgesamtheit, der die Stichprobe entnommen wurde
- Gültigkeit hängt davon ab, dass bei der Auswahl der Stichprobe keine Fehler begangen wurden.

**Zusammenhang von Grundgesamtheit (N) und Stichprobe (n)**



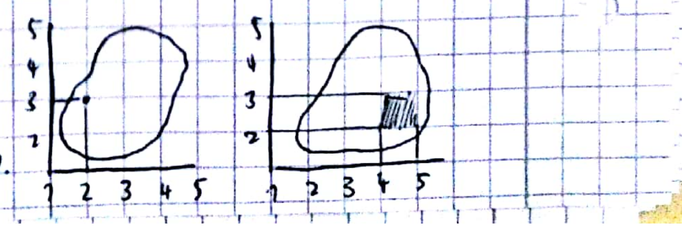

**Auswahlverfahren zur Ziehung einer Stichprobe**

zufällige Auswahl
Jedes Element der Grundgesamtheit hat die gleiche Chance, in die Stichprobe aufgenommen zu werden.
- reine Zufallsstichprobe
- systematische Stichprobe
- geschichtete Stichprobe
- Klumpenstichprobe

Nicht zufällige Auswahl
Jedes Element der Grundgesamtheit hat eine bestimmte Chance, in die Stichprobe aufgenommen zu werden.
- willkürliche Auswahl
- typische Auswahl
- Quotenauswahl
-

**Klumpenstichprobe (cluster sample)**
- Die Grundgesamtheit wird vor der Ziehung in räumliche Einheiten eingeteilt und dann alle Elemente einer zufällig gewählten Einheit ausgewählt
- Klumpen können definiert werden als natürliche Gruppierungen der Elemente einer Grundgesamtheit
  → Wahlbezirke → Wähler/innen
  → Schulklassen → Schüler/innen

**Stichprobenauswahl im Gelände**
Auswahl von Standorten an denen Bodenproben o. ä. entnommen werden sollen oder an denen Messgeräte ausgewählt werden.



## Geschichtete Zufallsauswahl

**Fragestellung:** Wie schätzt die deutsche Bevölkerung die Studienbedingungen an deutschen Unis ein? Eine Stichprobe mit proportionaler Schichtung würde 34 Studierende bei Tausend befragten ausmachen. (3,4% der deutschen sind Studenten (2,78 Mio von ~ 82.00 0000). Wie Aussagekräftig ist eine solche Stichprobe?

- geschichtete Zufallsauswahl soll sicherstellen, dass auch kl. Subpopulationen ('Minderheiten') in ausreichend großem Umfang in einer Stichprobe vertreten sind. Es gibt 2 typen proportional + disproportional

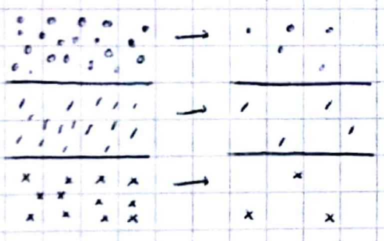
**Ausgangssituation:**
Grundgesamtheit (N) = 480
Stichprobe (n) = 60

**Schritt 1:** Einteilung der Elemente der Grundgesamtheit in Schichten (z.B. versch. Alter; oder Bildungsabschluss ..)

$240(\frac{1}{2})\ (480)$ $\quad 160(\frac{1}{3})(480)$ $\quad 80(\frac{1}{6})(480)$

**Schritt 2:** Zufallsstichprobe aus jeder einzelnen Schicht
proportional : (n = 60, davon 30, + 20 + 10)
disproportional : (n = 60, davon 20 + 20 + 20)

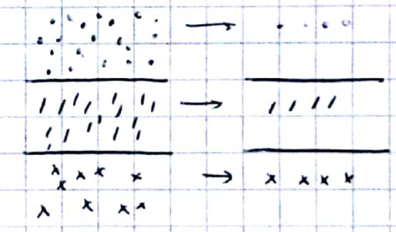
**Proportionale Schichtung**



Grundgesamtheit N hier eingeteilt in 3 Schichten (20, 16, 12) ; n = 12
proportional = 5 + 4 + 3 (= 12)

Der Anteil der Schichten in (n) entspricht dem Anteil der Schichten von (N)

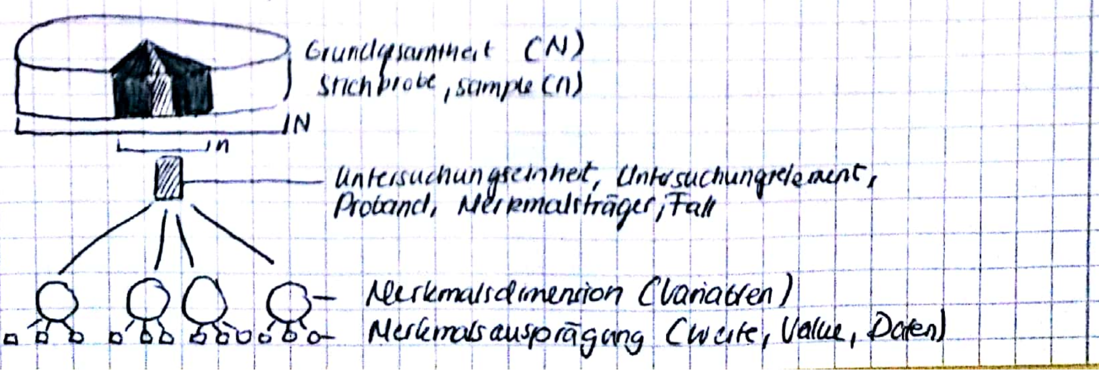
**Disproportionale Schichtung**



Grundgesamtheit mit 48 Elementen unterteilt in drei Schichten (20, 16, 12) n = 12
disproportional = ~~20 + 20 + 20~~ 4 + 4 + 4 (= 12)

Der Anteil der Schichten in [n] muss nicht dem Anteil der Schichten [N] entsprechen. Die Anteile für die Schichten in [n] sind frei wählbar.
Vor der Auswertung ist eine Gewichtung erforderlich, da Kriterium 'gleiche Chance in [n] zu gelangen' nicht gegeben ist.

**Elemente einer empirischen Datenbank**



## Merkmalsträger, Proband, Fall:

Diejenigen Elemente die untersucht werden (z.B. Probanden, die befragt werden, Bäume an denen man Flechten kartiert.) bezeichnet man je nach Art der Untersuchung als
- Untersuchungseinheit
- Untersuchungselemente
- Probanden, Merkmalsträger, Fälle

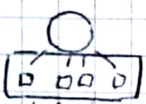
In der Regel werden sie durchnummeriert: $UE_1$, $UE_2$ ... $UE_n$, $PB_n$, $Fall_n$.

## Merkmalsdimensionen / Variablen:

Von den Untersuchungseinheiten erhebt man, je nach Ziel der geographischen Untersuchung, Daten (also Infos) zu einer bestimmten Zahl von Merkmalen (= Variablen) → Merkmalsdimension.

Merkmale eines Falles aus der Stichprobe z.B:
- Studienzahl
- Alter
- Semesterzahl

## Merkmalsausprägung (Value)



Diese Merkmalsdimensionen haben bei jeder Untersuchungseinheit (z.B bei jedem Probanden) einen bestimmten Wert (value) → Merkmalsausprägung.

Merkmalsausprägungen der Merkmalsdimensionen:
- Studienziel: BA, LA ...
- Alter: 22 J.; 23 J.;
- Semesterzahl: 2, 3.
- Familienstand: ledig; ...

## Eigenschaften von Variablen: (Merkmalsdimensionen)

**Quantitativ**
- Größenangabe
- kleiner / größer Relation
- Rechenoperation möglich
- Unterscheidung in diskret und stetig

| diskret | stetig |
|---|---|
| • (ganzzahlige) Variablen: kann nur bestimmte Werte annehmen → werden gezählt | • beliebige Variablen: kann in einem bestimmten Werte-Bereich jeden beliebigen Wert annehmen → werden gemessen |
| Bsp.: Anzahl der Kinder, Haushaltsgröße | Bsp.: Niederschlag, Fläche, Geschwindigkeit |

**Qualitativ**
- keine Größenangabe
- keine Rechenoperation möglich
- beschreiben einen Zustand, variieren nach Art oder Qualität

Bsp.: Nationalität, Konfession, Handelsklasse

Erhebung von Informationen zu Variablen sind mit unterschiedlichen Fragen möglich. Um Datenmengen zu strukturieren und mathematisch zu analysieren, werden empirisch ermittelten Daten Zahlen zugeordnet, die den jeweiligen (Merkmalsdimensionen / Values) entsprechen. → Kodierung. Wichtig: bereits die Formulierung einer Frage legt fest, welchem Datentyp sie erhalten bzw. welches Datenniveau vorliegt. Auf Grund ihrer Eigenschaften lassen sich Variablen einem oder mehreren Datenniveaus zuordnen.

## Datenniveaus (Mess- und Skalenniveaus)

- Daten, nominal skaliert (Nominalskala)
- Daten, ordinal skaliert (Ordinal-Skala)
- Daten, metrisch skaliert (metrische Skala: Intervall- und Rational-Skala)

## Skalenniveaus

| kategorial | | metrisch | |
|---|---|---|---|
| nominal | ordinal | intervall | rational |
| - nur einfache Begriffe<br>- nicht vergleichbar<br>- =, ≠ | - komparativ<br>- können geordnet werden<br>- =, ≠, >, < | - konstante Maßeinheiten<br>- äquidistant<br>- relativer Nullpunkt<br>- =, ≠, >, <, +, − | - konstante Maßeinheit<br>- äquidistant<br>- absoluter Nullpunkt<br>- =, ≠, >, <, +, −, x, ÷ |
| Bsp. Geschlecht, Wohnort | Schulart, Klasse | Temp (°C), Alter / Jahr (BC and AD) | (Temp °K), Alter / Jahr (BP), Größe |

### Nominales Skalenniveau (Nominal-Skala)

Unteres Messniveau

Merkmalsträgern einer Stichprobe von 'Studierenden' wird für die Variable 'Studiengang' zB (1) für Bachelor 100, (2) für Bachlor 50, (3) für Lehramt etc. zugeordnet. Anstatt der Bezeichnungen (der nominalen Werte) BA 100, BA 50, Lehramt werden den einzelnen Ausprägungen die Zahlen (1), (2), (3) etc. zugewiesen (Kodierung).

Statistische Analyse:

- absolute und relative Häufigkeiten, Modalwert

### Ordinales Datenniveau (Ordinal-Skala)

Mittleres Messniveau

Den Fällen der Stichprobe 'Studierende' werden nach ihrer Leistung die Zensuren / Noten 'sehr gut' (1), 'gut' (2) ... zugeordnet.

Mit der Zuweisung der 'Zensur'-Zahlen (1) bis (6) werden aber lediglich Aussagen über die Rangfolge möglich, da Unterschiede (Abstände) zwischen den Zensuren nicht gleich groß (äquidistant) sind.

Statistische Analyse:

- absolute und relative Häufigkeiten, Modalwert, Median
- Rangfolgen
- Nicht: arithmetisches Mittel

### Metrisches Datenniveau (Metrische Skala) – Intervallskala, Rationalskala

Oberes Messniveau

Sind Abstände zwischen den Merkmalsausprägungen (Werten / values) gleich groß (äquidistant), können exakte Abstände und Relationen zwischen den erhobenen Werten angegeben werden. Metrische Daten können arithmetrischen Operationen unterworfen werden; Angaben zu Relationen sind möglich, wenn Variable einen absoluten Nullpunkt hat.

- Intervall-Skala liegt vor, wenn eine Variable keinen absoluten Nullpunkt hat, zB. Temperatur in °C (30°C nicht doppelt so warm wie 15°C)
- Rational-Skala liegt vor, wenn Variable einen absoluten Nullpunkt hat, z.B. Entfernungen (100 km sind doppelt so lang wie 50 km)

### Warum muss nach Datenniveaus unterschieden werden?

Je nach Datenniveau sind nur bestimmte statistische Berechnungen zulässig / sinnvoll, nur bestimmte Interpretationen und Aussagen zulässig.

Bereits vor der Datenerhebung muss deshalb unbedingt klar sein, welches Niveau der Interpretation / Aussage im Bezug auf die Fragestellung angestrebt wird, damit ausreichend genau erhoben werden kann.

Wenn Sie z.B. für Aussagen zur Variable 'Entfernung zwischen Wohnort und Institut' Lagemaße wie arithmetisches Mittel und / oder Streuungsmaße wie Standardabweichung angeben möchten, müssen Sie diese Variable auf metrischem Datenniveau erfassen.

## Statistische Methoden in der Geographie - VL 5

20.11.17

## Lage- und Streuungsmaße

### Deskriptive Statistik:

- Dient der Beschreibung der Verteilung von Merkmalen, indem sie z.B. Häufigkeiten bestimmt, Durchschnittswerte bildet oder etwas über die Streuung eines Merkmals aussagt
- ist in der Lage große Datenmengen übersichtlicher zu machen, indem sie sie ordnet, gruppiert oder verdichtet.
- Erleichtert es, das Charakteristische, wichtige einer Verteilung herauszustellen.

### Lagemaße (Maße der zentralen Tendenzen)

- beschreiben das Zentrum (die Mitte) einer Verteilung, z.B.
  - → Modus (Modalwert) — ab nominalem Datenniveau möglich
  - → Median ($\tilde{x}$) — ab ordinalem Datenniveau möglich
  - → arithmetisches Mittel ($\bar{x}$) — ab metrischem Datenniveau möglich

### Streuungsmaße

- beschreiben die Lage der Einzelwerte um das Zentrum einer Verteilung -
  z.B. Spannweite = Differenz aus Maximum und Minimum

### Modalwert (Modus)

= häufigster Wert in einer Datenreihe
- kann für jedes Skalenniveau angegeben werden
- in manchen Fällen ist die Angabe des Modalwerts nicht sinnvoll

Wann ist die Angabe sinnvoll?
→ wenn in den Rohdaten häufig der gleiche Wert auftritt.
Bsp. Variable: Familienstand: v, v, l, v, g, v, v, l, l, v, v
→ Modalwert: v = verheiratet (tritt 7x auf, übrige insgesamt nur 4x)

Wann ist die Angabe nicht sinnvoll?
→ Wenn in den Rohdaten nur selten oder nie der gleiche Wert auftritt.
Bsp. Variable: Niederschlag: 122, 112, 133, 135, 125, 127, 128 ...
→ Modalwert: keiner (jeder Wert tritt nur 1x auf)

### Median

= mittlerer Wert einer aufsteigend geordneten Datenreihe
- bei ungeradem Stichproben-Umfang: 1.6  3.2  3.3  3.5  13.7

$$Me = x_{\frac{n+1}{2}} = \frac{5+1}{2} = 3 \rightarrow$$ also der dritte Wert der Datenreihe

- bei geradem Stichproben-Umfang:

Me = arithmetisches Mittel: $x_{\frac{n}{2}}$ = 1. Wert ; $x_{\frac{n}{2}+1}$ = 2. Wert

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Rohdaten: | 3,2 | 2,2 | 1,6 | 3,5 | 13,7 | 2,6 | 3,9 | 1,2 | 3,3 | 3,0 |
| Sortieren: | 1,2 | 1,6 | 2,2 | 2,6 | 3,0 | 3,2 | 3,3 | 3,5 | 3,9 | 13,7 |

Mittleres Element berechnen:
bei geradem SP-Umfang (n = 10)  3.1

Me = arith. Mittelwert aus Fall $x_{\frac{n}{2}} = \frac{10}{2} = 5$ und Fall $\frac{10}{2} + 1 = 6$
5. Wert + 6. Wert

arith. Mittel von $\frac{3,0 + 3,2}{2} = 3,1$

## Arithmetischer Mittelwert

$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ $\quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ $\quad \bar{x} = \frac{x_1 + x_2 + x_3 ... + x_n}{n}$ $\quad$ n (Elemente)

**Rohdaten:** 3.3; 3.2; 1.6; 3.5; 13.7; 2.6; 3.9; 1.2; 3.3; 3.0 → n = 10

**Berechnung:** $\frac{3.3 + 3.2 + 1.6 + 3.5 + 13.7 + 2.6 + 3.9 + 1.2 + 3.3 + 3.0}{10} = \underline{\underline{3,82}}$

## Median + arithmetisches Mittel → Vergleich

**Median**
- Rohdaten werden in aufsteigender Reihenfolge sortiert
- Median wird am mittleren Element abgelesen
- robust gegenüber Ausreißern und Extremwerten

Beispiel = 3.7

→ Median wird vom Extremwert 13.7 nicht beeinflusst

**Arithmetisches Mittel**
- alle Rohdaten gehen in Berechnungen ein (vgl. Formel):

$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

- anfällig gegenüber Ausreißern + Extremwerten

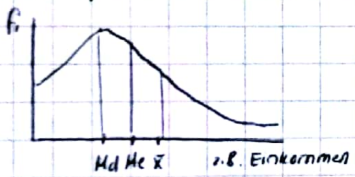
Beispiel = 3.82

→ Extremwert von 13.7 beeinflusst das arithmetische Mittel stark

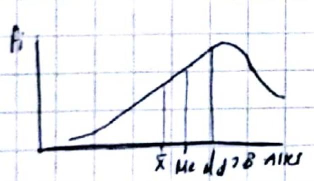
Aussagekraft von Lagemaßen (I)

## Aussagekraft von Lagemaßen (II)

$\bar{x} =$




positiv schief (rechts schief, links steil)
Ist der Modalwert kleiner als Median und arithmetisches Mittel, hat die Verteilung viele niedrige Werte und wenige hohe Werte.




negativ schief (links schief, rechts steil)
Ist der Modalwert größer als arithmetisches Mittel und Median, hat die Verteilung wenig niedrige Werte und viele hohe Werte.

## Gewichtetes Arithmetisches Mittel = 'gewogener Durchschnitt'

z.B.: Berechnung des monatlichen Durchschnittseinkommens in Städten mit unterschiedlichen Einwohnerzahlen.

| | $g_i$ - Einw. | $x_i$ Eink. |
|---|---|---|
| Stadt A | 500.000 | 1.500 |
| Stadt B | 100.000 | 2.800 |
| Stadt C | 250.000 | 1.900 |

Gewichtung der Einzelwerte mit einem bestimmten Faktor

$\bar{x}_g = \frac{\sum_{i=1}^{n} g_i x_i}{\sum_{i=1}^{n} g_i}$ $\quad \bar{x}_g = \frac{\bar{x}_1 \cdot n_1 + \bar{x}_2 \cdot n_2 ... + \bar{x}_k \cdot n_k}{n_1 + ... + n_k}$

g = Gewichtungsfaktor, z.B. Anzahl der Merkmalsträger, die die Merkmalausprägung x aufweisen.

$\frac{1{,}500 \cdot 500.000 + 2.800 \cdot 100.000 + 1.900 \cdot 250.000}{500.000 + 100.000 + 250.000} = 1770$

## Geometrisches Mittel →

## Geometrisches Mittel

Das geometrische Mittel wird für die Berechnung durchschnittlicher Wachstumsraten (Bsp: durchschnittliche Umsatzsteigerungen / Jahr) veränderungen der Bevölkerungszahlen / Jahr)

→ Wachstumsrate ist die prozentuale Veränderung gegenüber dem Vorjahr

$$GM = \sqrt[n]{\prod_{i=1}^{n} x_i} \qquad GM = \sqrt[n]{x_1 * x_2 * x_3 * \dots * x_n}$$

Bsp: Welchen durchschnittlichen Zinssatz erhalte ich bei 0%, 20% und 44% Zinssatz in drei Jahren.

$$GM = \sqrt[3]{1,2 * 1,0 * 1,44} = \sqrt[3]{1,728} = 1,2$$

## Harmonisches Mittel

Geeignetes Lagemaß für Größen die durch relativen Bezug auf eine Einheit definiert sind,
→ Geschwindigkeiten (Strecke pro Zeiteinheit).

$$HM = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \qquad HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}}$$

Bsp: 240 km Strecke, 1. 120 km = 40 kmh; 2. 120 km Strecke 120 km/h

(~~nicht 40 km/h + 120 km/h = 120 km/h / 2 = 8 km/h: Drei Stunden~~)

Aber: $\frac{120 \text{ km}}{40 \text{ km/h}} + \frac{120 \text{ km}}{120 \text{ km/h}} = 4h$ — $\boxed{\text{Zähler variable ist konstant}}$ = HM

$$\frac{240 \text{ km}}{4h} = 60 \text{ km/h}$$

Brüche addieren → gleicher Nenner!

$$HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2}} = \frac{2}{\frac{1}{40} + \frac{1}{120}} = 2x \frac{120}{4} = 60$$ (also 60 km/h im Schnitt)

Aber, bei 60 km/h + 120 km/h = 180 km/h / 2h = 90 km/h

## Arithmetisches Mittel bei klassifizierten Daten

| | | |
|---|---|---|
| 22 J = 2 Personen | 22 * 2 = | 44 |
| 23 J = 13 Personen | 23 * 13 = | 299 |
| 24 J = 11 Personen | 24 * 11 = | 264 |
| 25 J = 8 Personen | 25 * 8 = | 200 |
| 26 J = 6 Personen | 26 * 6 = | 156 |
| 27 J = 2 Personen | 27 * 2 = | 54 |
| Σ 42 | | |

1017 / 42 = 24,21 → arithmetisches Mittel

Klassenbreite = 2

| | | |
|---|---|---|
| 22-23 = 15 Pers. | 22,5 * 15 = | 337,5 |
| 24-25 = 19 Pers. | 24,5 * 19 = | 465,5 |
| 26-27 = 8 Pers. | 26,5 * 8 = | 212 |
| Σ 42 | | |

1015 / 42 = 24,17

Klassenbreite = 3

| | | |
|---|---|---|
| 22-24 = 26 Pers. | 23 * 26 = | 598 |
| 25-27 = 16 Pers. | 26 * 16 = | 416 |

1014 / 42 = 24,14

## Median bei klassifizierten Daten

| Einkommensklasse | Klassenmitte | Häufigkeiten absolute | und relative | (kum) Häufigkeite absolute | und relative |
|---|---|---|---|---|---|
| 1 | $x_i$ | $f_i$ | $p_i$ | $cf_i$ | $cp_i$ |
| 0 - 2000 | 1000 | 1 | 0,083 | 1 | 0,083 |
| 2000 - 4000 | 3000 | 6 | 0,500 | 7 | 0,583 |
| 4000 - 6000 | 5000 | 3 | 0,250 | 10 | 0,833 |
| 6000 - 8000 | 7000 | 1 | 0,083 | 11 | 0,916 |
| 8000 + mehr | (9000) | 1 | 0,083 | 12 | 0,999 |
| a | | Σ 12 | 0,999 | | |

Zunächst Modus (häufigster Wert in einer Zahlenreihe) → Medianintervall

Medianintervall - 2000 - 4000

$c_{mu}$ (exakte Untergrenze d. Medianintervalls) = 2000

$c_{mo}$ (exakte Obergrenze d. Medianintervalls) = 4000

$n$ (Σ Stichprobenumfang) = 12

$cf_u$ (kumulierte Häufigkeit unterhalb des Medianintervalls) = 1

$f_m$ (Häufigkeit im Intervall) = 6

$c_{mo} - c_{mu}$ (Breite des Medianintervalls) = 2000

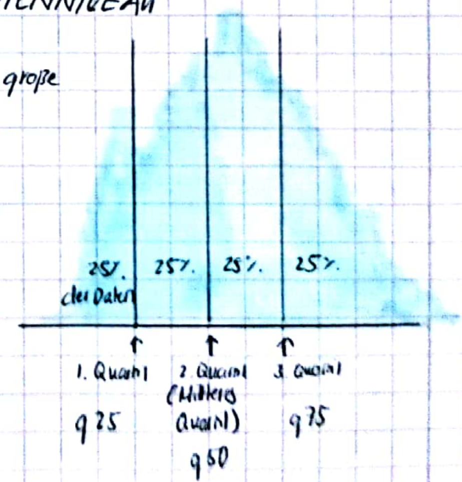
$$\text{Median} = c_{mu} + [(n/2 - cf_u) \cdot (c_{mo} - c_{mu}) / f_m]$$
$$Me = 2000 + [(12/2 - 1) \cdot 2000 / 6] = 3666{,}67 \text{ DM}$$

## VL 6 – Statistische Methoden in der Geographie

Streuungsmaße beschreiben die Lage der Einzelwerte um das Zentrum einer Verteilung.

- Quantile (Quartile, Quintile, Dezentile, Perzentile)
- Quartilsabstand, mittlerer Quartilsabstand
- Spannweite (range) = Differenz aus Maximum und Minimum (größtem und kleinstem Wert) einer Verteilung
- mittlere Abweichung, Varianz, Standardabweichung (absolute Streuungsmaße)
- Variationskoeffizient, relative Variabilität (relative Streuungsmaße)

Quartile (Sonderform der Quantile) ORDINALES DATENNIVEAU




- Quartile (Quintile, Perzentile) teilen Stichprobe in gleich große Gruppen (Quartile 4tel, Quintile 5tel, Perzentile 100tel)
- ermöglichen Aussagen zu einzelnen Abschnitten einer Stichprobe
- Elemente der Stichprobe werden aufsteigend geordnet (nach Größe)
- Median teilt Messwerte in obere und untere Teilste
- Unteres Quartil → Median der unteren Teilste
- Oberes Quartil → Median der oberen Teilste

→ 2 3 3 4 8 10 11 14 14 16 17 18 20 20 24

$Q_1$ (unter 4), $Me\,Q_2$ (unter 14), $Q_3$ (unter 18)

25% der Messwerte ≤ Quartilswert — 50% der Messwerte ≤ Quartilswert — 75% der Messwerte ≤ Quartilswert

## Berechnung von Quantilen, Quintilen und Perzentilen

- Quartil, 4 Teile ...
- Empirische Quantile teilen Daten einer Messreihe prozentual in 2 Teile
- mindestens $p * 100\%$ der Daten kleiner gleich dem Quantil ($p = 0{,}2 \rightarrow 20\%$)
- midestens $(1-p) * 100\%$ größer gleich (80%)

- Messdaten = geordnete Form: $x_1, x_2, x_3 \dots x_n$
- $0 < p < 1$

⇒ Berechnung des Quantils

$$\tilde{x} = \begin{cases} \frac{1}{2}(x_{n \cdot p} + x_{n \cdot p+1}), & \text{wenn } n \cdot p \text{ ganzzahlig} \\ x_{\lceil n \cdot p \rceil}, & \text{wenn } n \cdot p \text{ nicht ganzzahlig} \end{cases}$$

- Für reelle Zahl $x$ der Wert $\lceil x \rceil$ die kleinste Zahl, die größer gleich $x$ ist.

⇒ Aufrundungsfunktion

**Beispiel:** 2 3 3 4 8 10 11 14 14 16 17 18 20 20 24

$x_1\ x_2\ x_3\ \underline{x_4}\ x_5\ x_6\ x_7\ \underline{x_8}\ x_9\ x_{10}\ x_{11}\ \underline{x_{12}}\ x_{13}\ x_{14}\ x_{15}$

$p = 0{,}25$ unterstes Quantil; $p = 0{,}5$ Median; $p = 0{,}75$ oberstes Quantil

$\lceil n \cdot p \rceil = \lceil 15 \cdot 0{,}5 \rceil = \lceil 7{,}5 \rceil =$ aufgerundet $8 \rightarrow$ am 8. Wert ablesen (=14)

$\lceil n \cdot p \rceil = \lceil 15 \cdot 0{,}25 \rceil = \lceil 3{,}75 \rceil = 4 \rightarrow$ am 4. Wert ablesen (= 4)

$\lceil n \cdot p \rceil = \lceil 15 \cdot 0{,}75 \rceil = \lceil 11{,}25 \rceil = 12 \Rightarrow$ am 12. Wert ablesen (= 18)

$$\left[ \lceil n\, p \rceil = \lceil 14 \cdot 0{,}5 \rceil = 7 \rightarrow \frac{(14 : 0{,}5) + (14 \cdot 0{,}5 + 1)}{2} = \frac{7.\text{ Wert} + 8.\text{ Wert}}{2} = \frac{11 + 14}{2} = 12{,}5 \right]$$

## Berechnung von Perzentilen

- Perzentile, Hundertstelwerte, Prozentränge
- Verteilung wird in 100 gleiche Teile geteilt → 1%-Segmente
- Perzentile können als Quantile aufgefasst werden $100 \cdot p \rightarrow$ ganze Zahl
- Quantil $Q_{90} \triangleq$ Perzentil $P_{90}$: unterhalb dieses Punktes liegen 90% der Verteilung

$x_1 \dots x_n = (2..24)$, $p = 0{,}9$, $p = 0{,}4$

$\lceil n \cdot p \rceil = \lceil 15 \cdot 0.9 \rceil = \lceil 13{,}5 \rceil = 14 \rightarrow$ am 14. Wert ablesen $(= 20) \rightarrow$ 90% aller Werte dieser Verteilung sind kleiner gleich 20.

$\lceil n \cdot p \rceil = \lceil 15 : 0.4 \rceil = 6$, also $\frac{6. + 7.\text{ Wert}}{2} = \frac{10. + 11}{2} = 10{,}5 \rightarrow$ 40% aller Werte sind kleiner gleich 10,5.

## Quartilsabstand IQR (ORDINALES DATENNIVEAU)

- Abstand zwischen 1. und 3. Quartil
- $Q_3 - Q_1$ (bzw $Q_{75} - Q_{25}$) = IQR, QA
- $(Q_3 - Q_1)/2$ = mittlerer Quartilsabstand
- Hier $23 - 19 = 4$ Min
- Hier $(23-19)/2 = 2$ Min
- Je weniger die Werte eine Verteilung streuung
  - → desto kleiner der (mittlere) Quartilsabstand
  - → desto homogener die Merkmale der Stichprobe