

1: Datenerhebung und Häufigkeiten

Statistische Verfahren in der Geographie

Till Straube <straube@geo.uni-frankfurt.de>

Institut für Humangeographie
Goethe-Universität Frankfurt

1 Lernziele dieser Sitzung

Sie können...

- einige Grundbegriffe der Statistik definieren.
- Typen von Stichproben unterscheiden.
- Skalenniveaus von Variablen bestimmen.
- Häufigkeitsverteilungen beschreiben.

2 Statistische Praxis

Was ist Statistik? Je nach Perspektive kann Statistik vieles sein: ein Teilgebiet der Mathematik, ein Untersuchungsobjekt kritischer Forschung oder ein unbeliebtes Studienfach.

Im Rahmen dieser Veranstaltung soll Statistik als eine Zusammenstellung von Praktiken in der quantitativen Forschung verstanden werden, wobei ihre Anwendung stets im Mittelpunkt steht. Eine hilfreiche Definition findet sich bei Haseloff und Hoffmann (1968):

„Allgemein kann gesagt werden: Die Statistik hat es mit Zahlen zu tun, die entweder aus Abzählvorgängen oder aus Messungen gewonnen wurden. Ihre Aufgabe ist es, ein solches Zahlenmaterial in eine optimal übersichtliche und informationsreiche Form zu bringen, aus ihnen methodische Schlußfolgerungen zu ziehen und gegebenenfalls auch die Ursachen der analysierten Zahlenverhältnisse mit sachlichen Methoden aufzudecken.“ (Haseloff und Hoffmann 1968: 27)

2.1 Grundbegriffe der Statistik

2.1.1 Untersuchungselement

Untersuchungselemente (auch Untersuchungseinheiten, Merkmalsträger, bei Personen: Proband*innen, engl. *sampling unit*) sind die individuellen Gegenstände empirischer Untersuchungen. Bei einer Hochrechnung zur Bundestagswahl ist dies z.B. eine befragte Wählerin.

2.1.2 Stichprobe

Eine Stichprobe (engl. *sample*) ist die Menge aller Untersuchungselemente, deren Daten direkt erhoben werden. Die Anzahl der Untersuchungselemente in der Stichprobe wird in Formeln mit n bezeichnet. Bei einer Hochrechnung z.B. bilden alle tatsächlich befragten Wähler*innen die Stichprobe.

2.1.3 Grundgesamtheit

Die Grundgesamtheit (auch Population, engl. *population*) ist die Menge aller potentiell untersuchbaren Elemente, über die Aussagen getroffen werden sollen. Die Stichprobe ist eine Teilmenge der Grundgesamtheit. Die Anzahl der Elemente in der Grundgesamtheit wird in Formeln mit N bezeichnet. Bei einer Hochrechnung zur Bundestagswahl sind dies z.B. alle Wähler*innen (bzw. alle Wahlberechtigten, wenn Wahlbeteiligung von Interesse ist).

2.1.4 Variable

Variablen (auch Merkmale, engl. *variable*) sind Informationen über die Untersuchungselemente, die in einer Untersuchung von Interesse sind. Typischerweise unterscheiden sie sich von Untersuchungselement zu Untersuchungselement, sind also variabel. Bei einer Hochrechnung ist dies die Antwort auf die Frage: „Welche Partei haben Sie gerade gewählt?“

2.1.5 Wert

Ein Wert (auch Merkmalsausprägung, engl. *observation*) ist die erfasste Ausprägung einer Variable bei einem Untersuchungselement. In Formeln werden Werte mit $x_1, x_2, x_3, \dots, x_n$ durchnummeriert. Bei einer Hochrechnung kann die Variable „gewählte Partei“ für ein Untersuchungselement z.B. den Wert „CDU“ annehmen.

2.1.6 Kennwert

Kennwerte (auch Maßzahlen, Kennzahlen, engl. *summary statistics*) sind Zahlen, die aus den beobachteten Werten errechnet werden. Sie können beispielsweise Aufschluss über Mittelwerte und Verteilung einer Variable oder den Zusammenhang mehrerer Variablen geben. Bei einer Hochrechnung sind z.B. die relativen Häufigkeiten (in Prozent) der Variable „gewählte Partei“ von besonderem Interesse.

2.2 Taxonomien statistischer Verfahren

Statistische Verfahren werden in mehrerlei Hinsicht unterschieden, wie im Folgenden beschrieben. Dabei schließen sich verschiedene Kategorien nicht unbedingt aus, es gibt also durchaus statistische Verfahren, die z.B. als univariat *und* deskriptiv bezeichnet werden.

2.2.1 Uni-, bi- und multivariate Statistik

Bei diesen Bezeichnungen ist entscheidend, wie viele Variablen bei den jeweiligen Verfahren zum Einsatz kommen. Im Allgemeinen spricht man bei einer Variable von univariater Statistik, bei zwei Variablen von bivariater Statistik und bei mehr als zwei Variablen von multivariater Statistik. (Manchmal werden allerdings auch Verfahren mit nur zwei Variablen als multivariat bezeichnet.)

In dieser Veranstaltung beschäftigen wir uns zunächst mit univariaten, dann mit bivariaten Verfahren. Verfahren mit mehr als zwei Variablen werden nicht behandelt.

2.2.2 Deskriptive und schließende Statistik

Unabhängig von der Anzahl der Variablen unterscheidet man auch nach der Art und Weise des Vorgehens:

2.2.2.1 Deskriptive Statistik

Die deskriptive Statistik (auch: beschreibende Statistik) dient der Beschreibung der Verteilung von Merkmalen, indem sie z. B. Durchschnittswerte bildet, Häufigkeiten bestimmt oder etwas über die Streuung eines Merkmals aussagt. Sie kann so große Datenmengen übersichtlicher machen, indem sie diese ordnet, gruppiert oder verdichtet. Sie erleichtert es also, das Charakteristische, Wichtige zu erkennen.

2.2.2.2 Schließende Statistik

Die schließende Statistik (auch: analytische, operative Statistik, Inferenzstatistik, Prüfstatistik) verhilft dazu, von Eigenschaften einer Stichprobe auf Eigenschaften der Grundgesamtheit verallgemeinern bzw. schließen zu können (deshalb eben auch: schließende Statistik) und diese Einschätzung überprüfen zu können.

Die schließende Statistik wird weiter unterteilt in Schätz- und Teststatistik:

2.2.2.2.1 Schätzstatistik

Die Schätzstatistik schätzt Kennwerte der Grundgesamtheit aus den Kennwerten einer Stichprobe.

2.2.2.2.2 Teststatistik

Die Teststatistik überprüft, als wie wahrscheinlich oder unwahrscheinlich gemachte Schätzungen bzw. Hypothesen gelten können.

2.3 Ablauf einer statistischen Untersuchung

Eine typische Anwendung statistischer Verfahren in der Forschung folgt diesem Schema:

2.3.1 Datenerhebung

- Eigene Erhebung z.B. durch Zählen, Messen, Befragung (primärstatistische Daten)
 - Auswahl von Untersuchungseinheiten
 - Wahl der Datenniveaus
- Rückgriff auf vorhandenes Datenmaterial (sekundärstatistische Daten)

2.3.2 Datenaufbereitung

- Verdichtung des gewonnenen Datenmaterials und Digitalisierung in Form einer Datenmatrix
- Verschneidung von mehreren Datensätzen
- Vereinheitlichung und Säuberung der Daten
- Überblick verschaffen durch einfache Beschreibung von Häufigkeiten und Maßzahlen (deskriptive Statistik)

2.3.3 Datenauswertung

- Verdichtete Beschreibung von Verteilungsmustern einer Variable (univariate deskriptive Statistik)
- Verdichtete Beschreibung der Beziehung zwischen zwei Variablen (bivariate deskriptive Statistik)
- Schluss von Stichprobe auf Grundgesamtheit (Schätzstatistik)
- Testen von Hypothesen über die Grundgesamtheit (Teststatistik)

3 Grundlagen der Datenerhebung

3.1 Typen von Stichproben

3.1.1 Reine Zufallsstichprobe

Bei endlichen Grundgesamtheiten können Lotterieverfahren angewendet werden. Dabei wird allen Elementen der Grundgesamtheit eine Zahl zwischen 1 und N zugeordnet. Anschließend werden Zufallszahlen ausgewählt und die entsprechenden Elemente in die Stichprobe übernommen.

3.1.2 Systematische Zufallsstichprobe

Die Elemente einer endlichen Grundgesamtheit werden in eine Rangordnung gebracht (Nummerierung 1 bis N). Anschließend wählt man jedes (N/n) -te Element aus. So entsteht eine Stichprobe der Größe n .

3.1.3 Geschichtete Zufallsstichprobe

Die Elemente einer endlichen Grundgesamtheit werden in Schichten (Klassen) zusammengefasst. Anschließend zieht man eine Zufallsstichprobe aus jeder Schicht. Geschichtete Stichproben setzen die Kenntnis einiger Parameter der Grundgesamtheit voraus. Zur Aufteilung des Stichprobenumfangs auf die einzelnen Schichten wird in der Regel die proportionale Aufteilung gewählt.

3.1.4 Klumpenstichprobe

Hier ist die Grundgesamtheit schon in „natürliche“ Gruppen aufgeteilt (z.B. Schulklassen) und es werden mehrere dieser Gruppen (Klumpen, engl. *cluster*) nach einem Zufallsverfahren als Stichprobe gewählt.

„Man beachte, dass ein einzelner Klumpen (...) keine Klumpenstichprobe darstellt, sondern eine Ad-hoc-Stichprobe, bei der zufällige Auswahlkriterien praktisch keine Rolle spielen. Die Bezeichnung „Klumpenstichprobe“ ist nur zu rechtfertigen, wenn mehrere zufällig ausgewählte Klumpen vollständig untersucht werden.“ (Bortz und Schuster 2010: 81)

3.2 Variablentypen

3.2.1 Qualitative Variablen

Qualitative Variablen können nicht der Größe nach, sondern nur im Hinblick auf ihre Eigenschaft/Art („Qualität“) unterschieden werden (z.B. Parteizugehörigkeit, Telefonnummer, Automarke).

Qualitative Variablen, die nur zwei mögliche Werte annehmen können, nennt man „dichotome“ Variablen (etwa Antworten auf Ja-Nein-Fragen).

3.2.2 Quantitative Variablen

Quantitative Variablen können der Größe nach unterschieden werden (Bsp. Geburtenzahl, Arbeitslosenrate).

Quantitative Variablen können diskret oder stetig sein:

3.2.2.1 Diskrete Variablen

Diskrete Variablen (auch diskontinuierliche Variablen) können nur endlich viele, ganzzahlige Werte annehmen. Zwischen zwei Ausprägungen befindet sich eine abzählbare Menge anderer Ausprägungen (z.B. Anzahl eigener Kinder, Haushaltsgröße in Personen).

3.2.2.2 Stetige Variablen

Stetige Variablen (auch: kontinuierliche Variablen) können in einem bestimmten Bereich jede beliebige Ausprägung annehmen. Der Ausdehnungsbereich kennt keine Lücken, sondern ist als ein fortlaufendes Kontinuum vorstellbar: Bei stetigen Variablen können zwischen zwei Werten oder Ausprägungen unendlich viele weitere Ausprägungen oder Werte liegen (z.B. Körpergröße, Längengrad in Dezimalform).

3.3 Skalenniveaus

Eine Variable lässt sich aufgrund ihrer Eigenschaften einem Skalenniveau (auch Skalentyp, Messniveau, Datenniveau, engl. *level of measurement*) zuordnen. Bestimmte Rechenoperationen und statistische Verfahren setzen bestimmte Skalenniveaus voraus. Deshalb ist es wichtig zu wissen, welchem Skalenniveau eine Variable zuzuordnen ist.

Tabelle 1: Die vier wichtigsten Skalenniveaus.

Skalenart	Beispiel	Mögliche Aussagen	gültige Lagemaße
Nominalskala	Postleitzahl	Gleichheit, Verschiedenheit	Modus
Ordinalskala	Militärischer Rang	+ Größer-kleiner-Relationen	+ Median
Intervallskala	Temperatur in °C	+ Gleichheit von Differenzen	+ arithmetisches Mittel
Verhältnisskala	Körpergröße	+ Gleichheit von Verhältnissen	+ geometrisches Mittel

Variablen lassen sich immer auch einem niedrigeren Skalenniveau zuordnen. Dies geht allerdings mit Informationsverlust einher.

Die im Folgenden beschriebenen Skalenniveaus sind nicht deckungsgleich mit den o.g. Variablentypen. Intervall- und Verhältnisskalen können z.B. jeweils diskret oder stetig sein.

In Tabelle 1 sind die wichtigsten Skalenniveaus im Überblick aufgeführt. „Gültige Lagemaße“ sind dabei als Zusatzinformation aufgelistet und werden erst in der nächsten Sitzung behandelt.

3.3.1 Nominalskala

Die Merkmalsausprägungen einer Variable stehen je „für sich“; sie lassen sich nicht sinnvoll in eine Rangordnung bringen oder gar miteinander verrechnen.

Die einzige Aussage, die sich über zwei Werte in einer Nominalskala treffen lässt, ist dass sie gleich oder nicht gleich sind.

Beispiele: Postleitzahlen, Telefonnummern, Staatsangehörigkeit, Krankheitsklassifikationen

3.3.2 Ordinalskala

Die Merkmalsausprägungen einer Variablen lassen sich sinnvoll in eine Rangordnung bringen, die Abstände zwischen den Merkmalsausprägungen aber lassen sich nicht sinnvoll quantifizieren.

Über zwei Werte in einer Ordinalskala lässt sich nicht nur sagen, ob sie gleich oder verschieden sind (wie in der Nominalskala), sondern darüber hinaus, welcher Wert bei Verschiedenheit größer ist.

Beispiele: Militärische Ränge, Windstärken, pauschale Häufigkeitsangaben (sehr oft ... nie), Zufriedenheitsangaben (sehr zufrieden ... unzufrieden)

3.3.3 Metrische Skalen (oder Kardinalskalen)

Abstände zwischen den Merkmalsausprägungen lassen sich exakt angeben.

Zusätzlich zu den Möglichkeiten der Ordinalskala können auf einer metrischen Skala Rechenoperationen auch sinnvoll auf die Differenzen zwischen den Merkmalsausprägungen angewendet werden.

Metrische Skalen werden unterteilt in Intervall- und Verhältnisskalen:

3.3.3.1 Intervallskala

Maßeinheit und Wahl des Nullpunktes sind willkürlich gewählt.

Beispiele: Grad Celsius, Geburtsjahr als Jahreszahl („1961“), in der Praxis häufig: subjektive Bewertung auf einer Skala von 1 bis 10.

3.3.3.2 Verhältnisskala (auch Ratioskala)

Es gibt einen invarianten (absoluten, natürlichen) Nullpunkt.

In einer Verhältnisskala lassen sich über alle o.a. Möglichkeiten hinaus auch Aussagen über Verhältnisse zwischen Werten treffen (z.B. „ x_1 ist doppelt so groß wie x_2 “).

Beispiele: Lebensalter in Jahren, Haushaltsgröße, Körpergröße, Körpergewicht

4 Häufigkeitsverteilungen

4.1 Urliste

Die Urliste ist eine ungeordnete Liste aller erfassten Werte.

Für die statistische Erhebung „Anfangsbuchstaben der Vornamen von Teilnehmenden an einer Statistikvorlesung“ könnte die Urliste z.B. so aussehen:

T J D T E N D F F M A J V T T V A L V P J K P M F M A J N A C I T P B A P H T L
N S P C K J K L J R E Y M K H M N L A A L L M L J G P L B F L J J V M P C J M J
S A M M M P A A L L O C J L P L V F J R M A V K S B B B N C A A T J P C F L E B
L C A K A L T V Y P F L J S T T N R J A S E L M L T A E B M N M V D P P L N L B
A A J M L N N S H M

4.2 Geordnete Liste

Die geordnete Liste bringt die Werte der Urliste in eine geeignete Reihenfolge, so dass die unterschiedlichen Werte leicht gezählt werden können:

A A A A A A A A A A A A A A A A B B B B B B B C C C C C C D D D E E E
E E F F F F F F G H H H I J J J J J J J J J J J J J J J K K K K K L L L
L L L L L L L L L L L L L L L L L M M M M M M M M M M M M M N N N N
N N N N N O P P P P P P P P P P P R R R S S S S S S T T T T T T T T T T
V V V V V V V Y Y

4.3 Häufigkeiten

Die absoluten Häufigkeiten erhält man durch einfaches Abzählen der jeweiligen Werte. Für die relativen Häufigkeiten teilt man diese Zahl durch n . Kumulierte Häufigkeiten zählen die bisherigen Summen bzw. Anteile zusammen (s. Tabelle 2).

Tabelle 2: Tabelle mit kumulierten Häufigkeiten

Buchstabe	Absolute Häufigkeit f	f_{kum}	Relative Häufigkeit	$\%_{kum}$
A	19	19	11,2%	11,2%
B	8	27	4,7%	15,9%
C	7	34	4,1%	20%
D	3	37	1,8%	21,8%
E	5	42	2,9%	24,7%
F	7	49	4,1%	28,8%
G	1	50	0,6%	29,4%
H	3	53	1,8%	31,2%
I	1	54	0,6%	31,8%
J	17	71	10%	41,8%
K	6	77	3,5%	45,3%
L	22	99	12,9%	58,2%
M	17	116	10%	68,2%
N	10	126	5,9%	74,1%
O	1	127	0,6%	74,7%
P	13	140	7,6%	82,4%
R	3	143	1,8%	84,1%
S	6	149	3,5%	87,6%
T	11	160	6,5%	94,1%
V	8	168	4,7%	98,8%
Y	2	170	1,2%	100%

Softwarehinweis

In R lässt sich mit dem Befehl `table()` eine einfache Häufigkeitstabelle aus Rohdaten erstellen.

4.4 Stabdiagramme

Die so ermittelten Häufigkeiten lassen sich als Stabdiagramm (auch Säulen-, Streifen-, Balkendiagramm, engl. *bar chart*) darstellen (s. Abbildung 1).

Softwarehinweis

In R lautet der Standardbefehl zur Erstellung eines Stabdiagramms `barplot()`.

4.5 Quantitative Variablen

Das oben beschriebene Verfahren funktioniert gut für qualitative Variablen (und diskrete Variablen mit wenigen unterschiedlichen Werten). Für quantitative Variablen wird ein anderes Verfahren empfohlen.

Zur Veranschaulichung soll diese geordnete Liste von Messwerten des Stammdurchmessers von Schwarzkirschen (Beispieldatensatz `trees` aus R Core Team 2017) dienen:

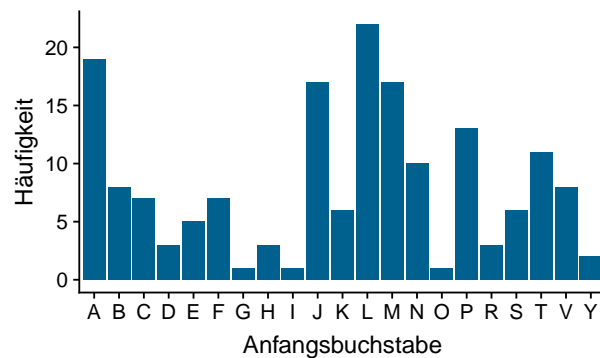


Abbildung 1: Stabdiagramm

Tabelle 3: Häufigkeitstabelle mit klassierten Werten

Durchmesser	Absolute Häufigkeit f	f_{kum}	Relative Häufigkeit	$\%_{kum}$
über 8 bis 10 Zoll	3	3	9,7%	9,7%
über 10 bis 12 Zoll	12	15	38,7%	48,4%
über 12 bis 14 Zoll	6	21	19,4%	67,7%
über 14 bis 16 Zoll	3	24	9,7%	77,4%
über 16 bis 18 Zoll	6	30	19,4%	96,8%
über 18 bis 20 Zoll	0	30	0%	96,8%
über 20 bis 22 Zoll	1	31	3,2%	100%

8,3 8,6 8,8 10,5 10,7 10,8 11,0 11,0 11,1 11,2 11,3 11,4 11,4 11,7 12,0 12,9
12,9 13,3 13,7 13,8 14,0 14,2 14,5 16,0 16,3 17,3 17,5 17,9 18,0 18,0 20,6

Für solche Verteilungen müssen zuerst Klassen (engl. *bins*) gebildet werden, in denen die Werte dann zusammengefasst werden (s. Tabelle 3).

Für die Wahl der Klassengrenzen gibt es zwei feste Regeln:

- Alle Werte müssen abgedeckt sein.
- Die Klassen dürfen sich nicht überlappen.

Zusätzlich sollten folgende Konventionen nach Möglichkeit befolgt werden:

- Klassen sollten gleich große Wertebereiche abdecken.
- Alle Klassen sollten besetzt sein.
- Klassengrenzen sollten möglichst glatte Zahlen sein.
- Aus Gründen der Übersichtlichkeit sollten nicht mehr als 20 Klassen gewählt werden.
- Klassengrenzen sollten „Klumpen“ mit ähnlichen Werten nicht trennen.

Die Darstellung erfolgt in so genannten Histogrammen (engl. *histogram*). Abbildung 2 enthält ein Beispiel für ein Histogramm.

Softwarehinweis

In R können Histogramme mit `hist()` erstellt werden.

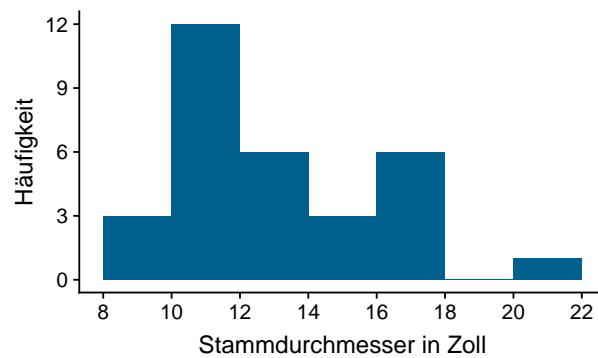


Abbildung 2: Histogramm

4.6 Polygone

Statt ausgefüllten Flächen wie im Histogramm lassen sich für die Häufigkeiten auch Punkte setzen, die dann mit Linien verbunden werden. So entsteht ein Häufigkeitspolygon (s. Abbildung 3).

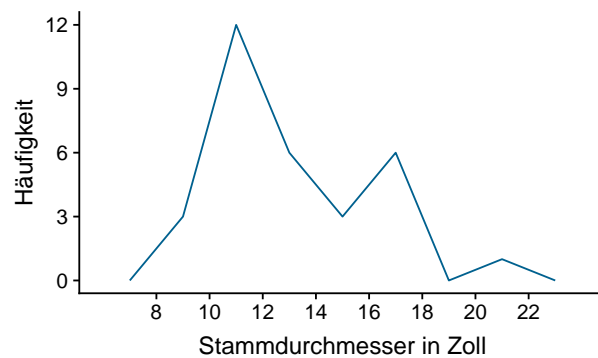


Abbildung 3: Polygonzug

4.7 Eigenschaften von Häufigkeitsverteilungen

Polygone von Häufigkeitsverteilungen (insbesondere in geglätteter Form) ergeben Annäherungen an so genannte Dichtefunktionen (engl. *density functions*). Diese lassen sich mit Attributen (uni-/bimodal, schmal-/breitgipflig, etc.) beschreiben, wie in Abbildung 4 veranschaulicht.

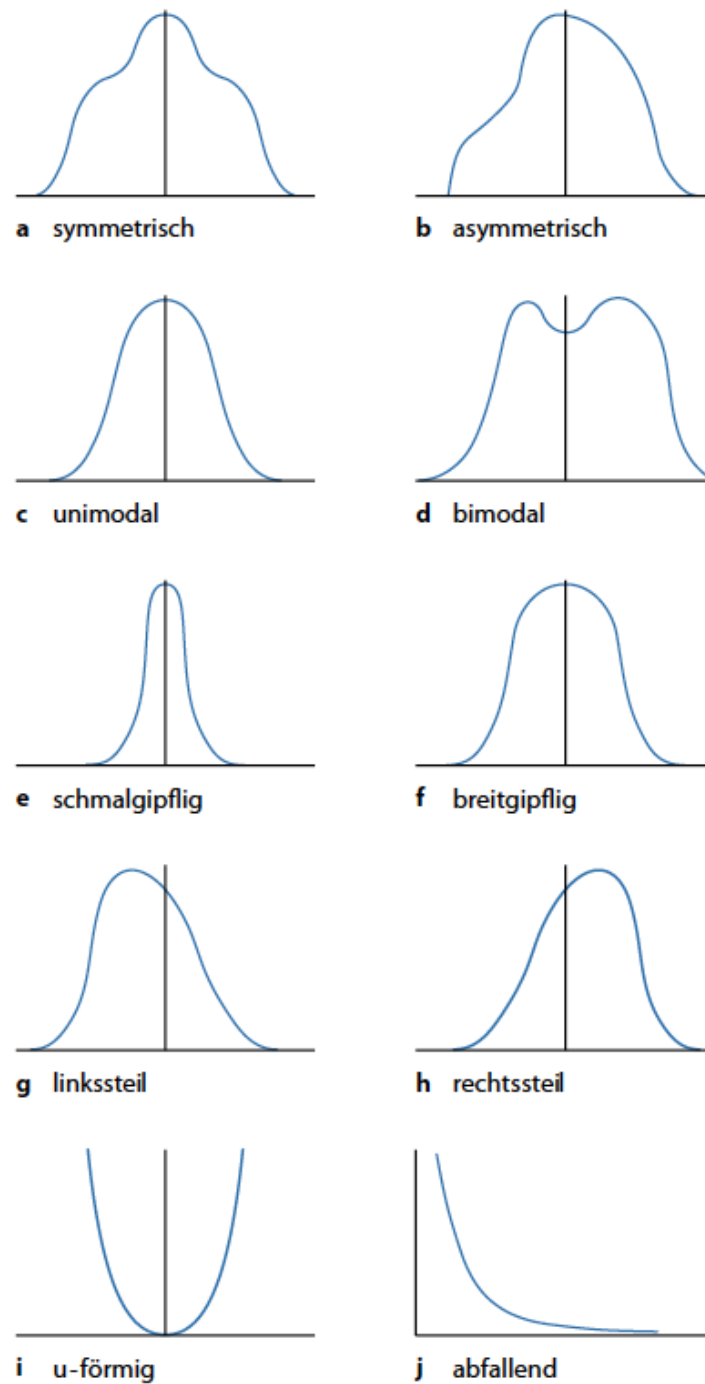


Abbildung 4: Merkmale von Verteilungen (aus: Bortz und Schuster 2010: 42)

5 Aufgaben

Die folgenden Aufgaben sind zur eigenständigen Überprüfung Ihrer Lernleistung gedacht (als Vor- oder Nachbereitung der Vorlesung, oder als Klausurübung) und nicht etwa als Hausaufgabe.

5.1 Aufgabe 1

Bestimmen Sie das Skalenniveau der folgenden Variablen. Kennzeichnen Sie darüber hinaus, ob die Variable qualitativ, diskret oder stetig ist.

- a) Lebensalter in Jahren
- b) Regenmenge in mm
- c) Güteklasse
- d) Passagieraufkommen
- e) Baujahr
- f) Geschwindigkeit in km/h
- g) Sozialstatus (Unter-, Mittel und Oberschicht)
- h) Temperatur in °F
- i) Fläche eines Bundeslands in km²
- j) Temperatur in K
- k) Einwohnerzahl
- l) Pegelstand
- m) Staatsangehörigkeit
- n) Interesse an Statistik (gering bis hoch)
- o) Klausurnote
- p) Bodentyp
- q) Entfernung zum Stadtzentrum in km
- r) Körpergröße
- s) Kleidergröße (S bis XXL)
- t) Monatliches Nettoeinkommen

5.2 Aufgabe 2

Folgende Werte seien erfasst über die Lebensdauer von Klimaanlage in Stunden (Beispieldatensatz `aircondit7` aus R Core Team 2017):

14 23 15 139 13 39 188 22 50 3 36 46 30 5 102 5 88 22 197 72 210 97 79 44

- a) Erstellen Sie eine Häufigkeitstabelle. Welche Klassen wählen Sie und warum?
- b) Zeichnen Sie ein Histogramm.
- c) Beschreiben Sie die Verteilung.

6 Tipps zur Vertiefung

6.1 Grundbegriffe

- YouTube-Kanal „Kurzes Tutorium Statistik“: [Statistische Grundbegriffe](#)

- Kapitel 1.1 in Bortz und Schuster (2010)
- Kapitel 1.1 in Benninghaus (2007)
- Kapitel 2.1 in Bahrenberg, Giese und Nipper (2010)

6.2 Stichproben

- Kapitel 6.1 in Bortz und Schuster (2010)
- Kapitel 2.3 in Bahrenberg, Giese und Nipper (2010)

6.3 Skalenniveaus

- Kapitel 1.2 in Bortz und Schuster (2010)
- Kapitel 2.1 in Benninghaus (2007)
- Kapitel 2.2 in Bahrenberg, Giese und Nipper (2010)
- YouTube-Kanal „Kurzes Tutorium Statistik“: [Skalenniveaus](#)

6.4 Häufigkeiten und Diagramme

- YouTube-Kanal „Kurzes Tutorium Statistik“: [Stabdiagramme und Histogramme](#)
- Kapitel 3.1 und 3.2 in Bortz und Schuster (2010)
- Kapitel 1.2 in Benninghaus (2007)
- Kapitel 4.1 in Bahrenberg, Giese und Nipper (2010)

Quellen

Bahrenberg, Gerhard, Ernst Giese und Josef Nipper. 2010. *Statistische Methoden in der Geographie*. 5. Aufl. Bd. 1. Univariate und bivariate Statistik. Stuttgart: Bornträger.

Benninghaus, Hans. 2007. *Deskriptive Statistik. Eine Einführung für Sozialwissenschaftler*. 11. Aufl. Wiesbaden: VS Verlag.

Bortz, Jürgen und Christof Schuster. 2010. *Statistik für Human- und Sozialwissenschaftler*. 7. Aufl. Berlin: Springer.

Haseloff, Otto W. und Hans-Joachim Hoffmann. 1968. *Kleines Lehrbuch der Statistik*. Berlin: de Gruyter.

R Core Team. 2017. R: A Language and Environment for Statistical Computing. <https://www.R-project.org/>.