



Technische Hochschule
Ingolstadt

Angewandte Mathematik - Statistik

– Vorlesung –

Wintersemester 2024/25

Dozent: Prof. Dr. Sören Gröttrup

Vorlage von Prof. Dr. Max Krüger

Fakultät Informatik, Technische Hochschule Ingolstadt (THI)



Prof. Dr. Sören Gröttrup

Maschinelles Lernen und Statistik

Fakultät für Informatik

E-Mail: Soeren.Groettrup@thi.de

Raum: A229

Sprechstunde: Donnerstags, 11:30 -12:30 Uhr

Moodle-Kurs: <https://moodle.thi.de/course/view.php?id=9265>

(gleicher Raum wie bei Prof. Krüger)

Der vorliegende Foliensatz ist ausschließlich für den persönlichen, vorlesungsinternen Gebrauch im Rahmen der Vorlesungen zur Mathematik und Statistik für anwendungsorientierte Informatik an der Fakultät Informatik der Technischen Hochschule Ingolstadt (THI) bestimmt.

Der Foliensatz wird kontinuierlich korrigiert, aktualisiert und erweitert.

– Urheberrechtlich geschütztes Material –

Die Weitergabe an Dritte sowie Veröffentlichungen in jeglicher Form (insb. Hochladen ins Internet, Social Media, Videoplattformen, etc.) sind u.a. aus urheberrechtlichen Gründen in keinem Fall gestattet.



Stochastik

Ist ein Teilgebiet der angewandten Mathematik, das aus Wahrscheinlichkeitstheorie und (mathematischer) Statistik besteht und die Beschreibung, Modellierung und Auswertung zufälliger Ereignisse und Phänomene behandelt.

Wahrscheinlichkeitstheorie

Die mathematische Disziplin zur Untersuchung und Beschreibung von Zufall.

Definition (Wikipedia)

Statistik „ist die Lehre von Methoden zum Umgang mit quantitativen Informationen“ (Daten). Sie ist eine Möglichkeit, „eine systematische Verbindung zwischen Erfahrung (Empirie) und Theorie herzustellen“. Unter Statistik versteht man die Zusammenfassung bestimmter Methoden zur Analyse empirischer Daten.

Aufgaben der Statistik

Aufbereitung und Analyse von quantitativen Informationen (Daten / Merkmalen) mit dem Ziel der Informationsgewinnung. Das beinhaltet:

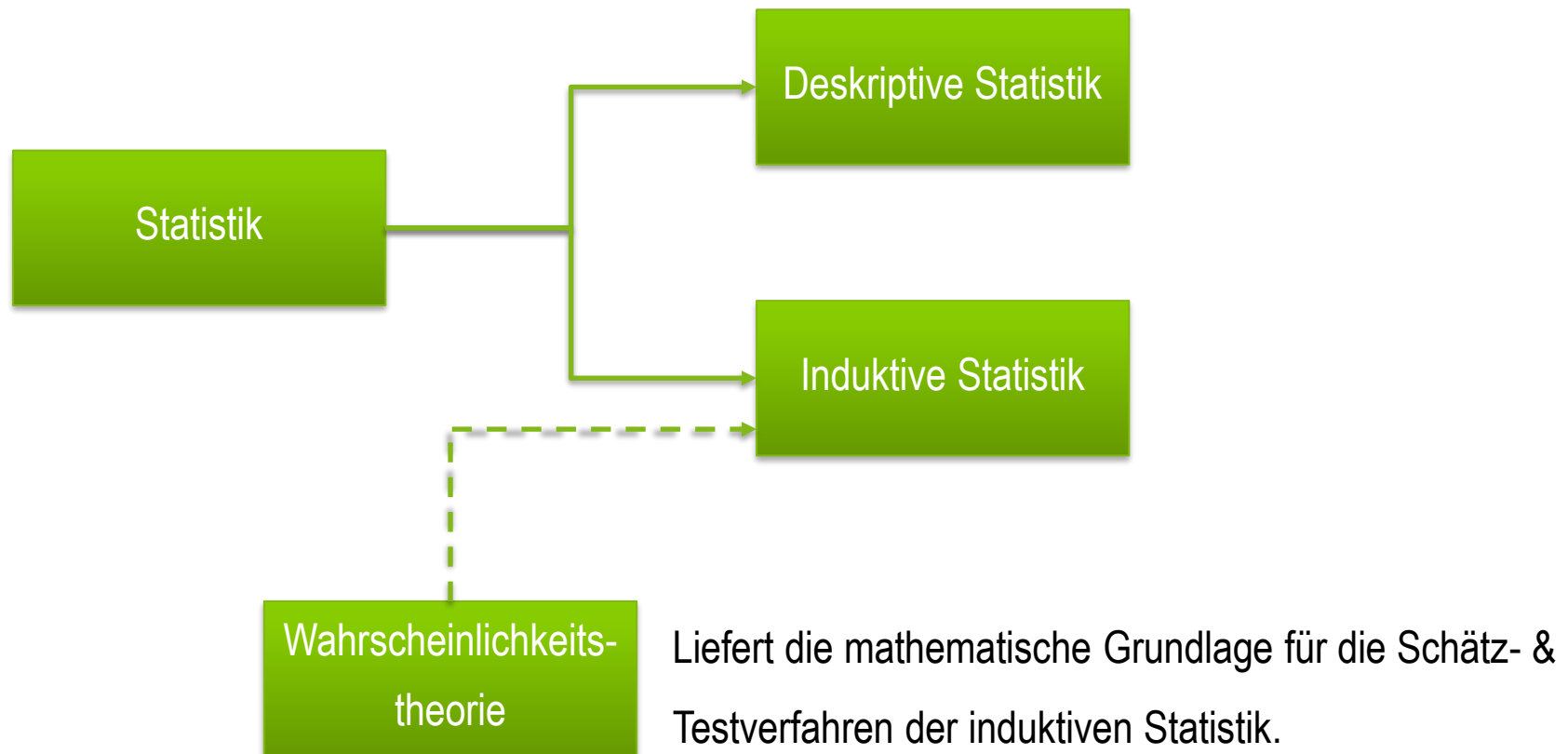
- Daten sammeln
- Daten darstellen und analysieren
- Daten interpretieren
- Prognosen und Entscheidungen treffen

Deskriptive Statistik (*beschreibende Statistik*)

Dient der beschreibenden und grafischen Aufbereitung sowie Komprimierung von Daten, insbesondere zur Präsentation von umfangreichem Datenmaterial.

Induktive Statistik (*schließende bzw. inferenzielle Statistik*)

Stellt Methoden bereit, um ausgehend von Daten einer (Zufalls-)Stichprobe Aussagen über die Grundgesamtheit treffen zu können (statistische Schlüsse ziehen).



Deskriptive Statistik

- VE 01: Merkmale und Stichproben
- VE 02: Tabellarische und grafische Darstellungen von Daten
- VE 03: Lage- und Streuungsmaße
- VE 04: Korrelation und Regressionsanalyse

Deskriptive Statistik

Zufall und Wahrscheinlichkeit

- VE 05: Kombinatorik, Zufallsvorgänge, Ereignisse und Wahrscheinlichkeiten
- VE 06: Wahrscheinlichkeitsrechnung und bedingte Wahrscheinlichkeiten

Zufallsvariablen und Verteilungen

- VE 07: Zufallsvariablen, Erwartungswerte, Varianzen und Verteilungsfunktionen
- VE 08: Wahrscheinlichkeitsverteilungen

*Wahrscheinlichkeits-
theorie*

Schätztheorie

- VE 09: Grenzwertsätze, Schätzfunktionen und Quantile
- VE 10: Konfidenzintervalle

Testtheorie

- VE 11: Hypothesentests
- VE 12: Parameter-, Anpassungs- und Unabhängigkeitstests

Induktive Statistik



Statistik

Themengebiet: Deskriptive Statistik
VE 01: Merkmale und Stichproben

Wintersemester 2024/25

Dozent: Prof. Dr. Sören Gröttrup

Vorlage von Prof. Dr. Max Krüger

Fakultät Informatik, Technische Hochschule Ingolstadt (THI)

Der vorliegende Foliensatz ist ausschließlich für den persönlichen, vorlesungsinternen Gebrauch im Rahmen der Vorlesungen zur Mathematik und Statistik für anwendungsorientierte Informatik an der Fakultät Informatik der Technischen Hochschule Ingolstadt (THI) bestimmt.

Der Foliensatz wird kontinuierlich korrigiert, aktualisiert und erweitert.

– Urheberrechtlich geschütztes Material –

Die Weitergabe an Dritte sowie Veröffentlichungen in jeglicher Form (insb. Hochladen ins Internet, Social Media, Videoplattformen, etc.) sind u.a. aus urheberrechtlichen Gründen in keinem Fall gestattet.

Thema: Merkmale und Stichproben

- Grundgesamtheit und statistische Einheiten
- Merkmale und Merkmalsausprägungen
- Qualitative und quantitative Merkmale
- Diskrete und stetige Merkmale
- Nominalskala, Ordinalskala und metrische Skala
- Intervallskala, Verhältnisskala und Absolutskala
- Rangordnung der Skalenarten
- Vollerhebung, Teilerhebung und Stichprobe
- Stichprobenauswahlverfahren

VE 01: Merkmale und Stichproben

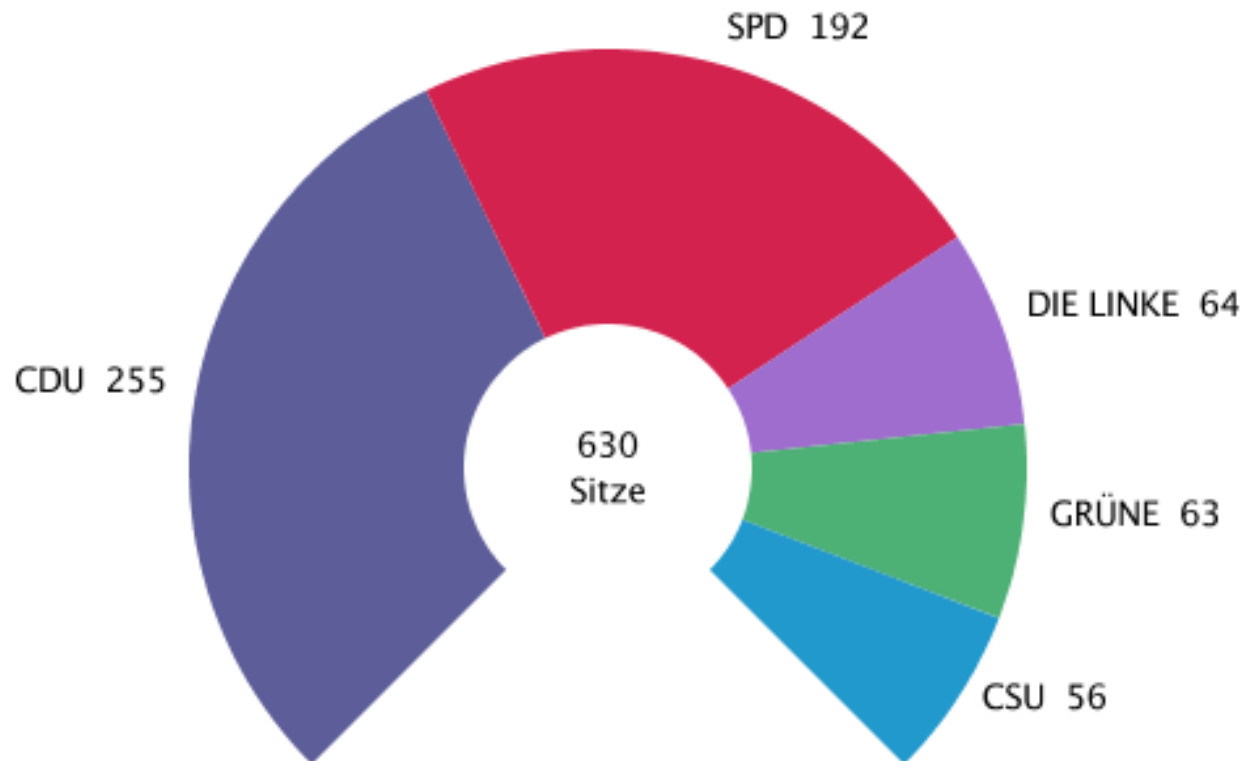
1.1 Einstieg: Bundestagswahl 2013

1.2 Grundgesamtheit und Merkmale

1.3 Stichproben und Stichprobenauswahlverfahren

Sitzverteilung nach vorläufigem Ergebnis der Bundestagswahl 2013:

Sitzverteilung



Quelle: Der Bundeswahlleiter, http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/ergebnisse/bundesergebnisse/grafik_sitze_99.html/ (abgerufen: 30.09.2013)

Stimmanteile nach vorläufigem Ergebnis der Bundestagswahl 2013:

Wahlberechtigte: 61 903 903

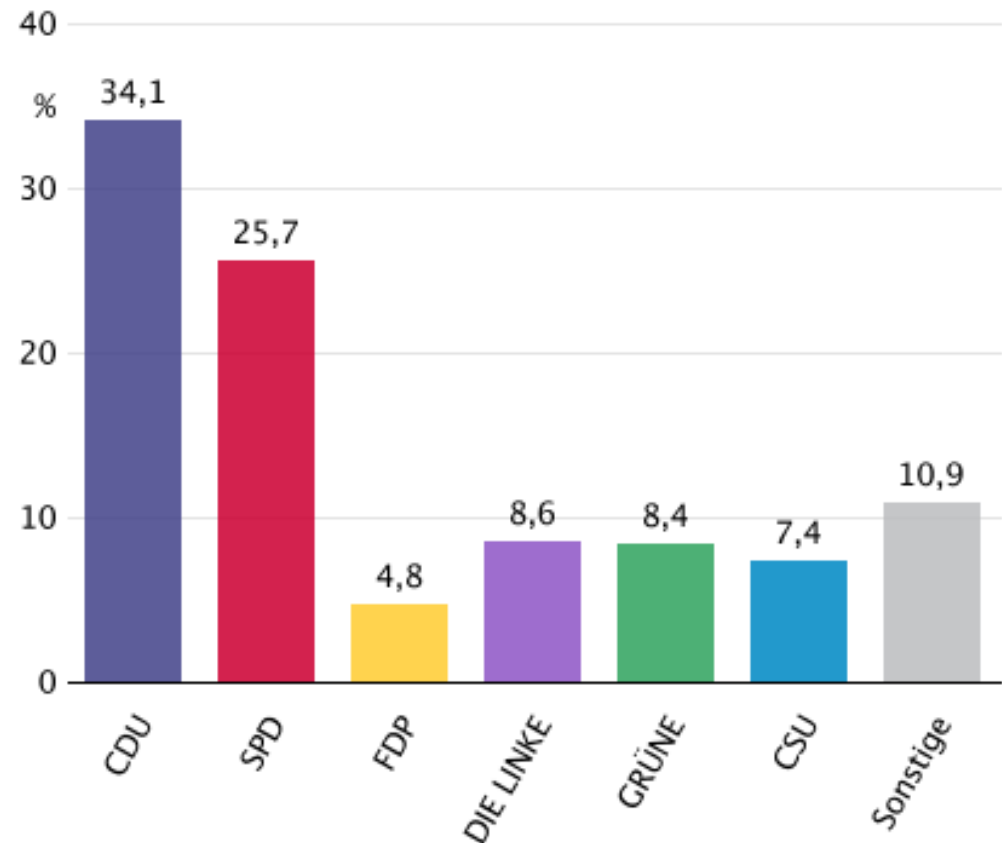
Wahlbeteiligung: 71.5 %

Ungültige Stimmen: 1.3 %

Aufgabe:

Wieviel Prozent der Wahlberechtigten haben für die CDU/CSU - Fraktion gestimmt?

Stimmenanteile



Quelle: Der Bundeswahlleiter, http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/ergebnisse/bundesergebnisse/index.html (abgerufen: 30.09.2013) und http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/ergebnisse/bundesergebnisse/grafik_stimmenanteile_99-2.html/ (abgerufen: 30.09.2013)

VE 01: Merkmale und Stichproben

1.1 Einstieg: Bundestagswahl 2013

1.2 Grundgesamtheit und Merkmale

1.3 Stichproben und Stichprobenauswahlverfahren

Grundgesamtheit (Population)

Die Menge von räumlich und zeitlich eindeutig definierten Objekten einer statistischen Untersuchung, die hinsichtlich bestimmter – vom Ziel der Untersuchung abhängender – Kriterien übereinstimmen.

Statistische Einheiten (***Merkmalsträger**, Untersuchungseinheiten, Messobjekte*)

Die Elemente der Grundgesamtheit, d.h. Personen oder Objekte, deren Eigenschaften für die bestimmte Untersuchung von Interesse sind.

Beispiele für Grundgesamtheiten:

- alle Bundesbürger
- alle mittelständischen Unternehmen

Merkmal

Eine spezielle Eigenschaft statistischer Einheiten, die in Hinblick auf das Ziel der konkreten statistischen Untersuchung von Interesse ist. Die Begriffe *Merkmal* und *Variable* werden etwas ungenau häufig synonym verwendet.

Merkmalausprägungen

Die möglichen Werte, die ein Merkmal annehmen kann.

Die Menge aller möglichen Merkmalausprägungen heißt **Wertebereich** des Merkmals.

Beispiele:

- Geschlecht, Alter, etc. sind Merkmale von Bundesbürgern.
- Weiblich, männlich sind Ausprägungen des Merkmals Geschlecht.
- 0,1,2,3,... 100 sind Ausprägungen des Merkmals Alter

Statistische Untersuchung

Ziel einer **statistischen Untersuchung** ist es, Aussagen über die Grundgesamtheit anhand der beobachteten Ausprägungen (Daten) der zu untersuchenden Eigenschaften der statistischen Einheiten zu treffen.

Stichprobe

Eine Teilmenge der Grundgesamtheit, für die diese Daten ermittelt werden. Aus verschiedenen Gründen wird dabei selten die vollständige Grundgesamtheit betrachtet (Kosten- und Zeitaufwand, zerstörende Prüfung, etc.)

Multivariate (mehrdimensionales) Merkmale

Sie entstehen durch Kombination **univariater** (einzelner) Merkmale einer Einheit.

Anmerkung: Ein zweidimensionales Merkmal heißt auch **bivariat**.

Beispiel:

Die Kombination von Alter und Geschlecht zu einem kombinierten Merkmal bildet ein bivariates Merkmal.

Aufgabe:

Überlegen Sie sich eine Fragestellung bezüglich eines bivariaten Merkmals, die bei Zerlegung in zwei univariate Merkmale nicht beantwortet werden kann.



Qualitative Merkmale

Sind solche Eigenschaften von statistischen Einheiten, die **qualitativ**, d.h. **der Beschaffenheit nach, artmäßig** variieren. Sie besitzen (typischerweise) nur endlich viele Ausprägungen.

- Beispiele: Geschlecht, Religion, Rechtsform von Unternehmungen.

Quantitative Merkmale

Sind solche Eigenschaften von statistischen Einheiten, die **quantitativ**, d.h. **der Größe nach oder zahlenmäßig**, variieren. Ihre Merkmalsausprägungen sind von vornherein Zahlen, mit oder ohne Maßeinheit.

- Beispiele: Alter, Kinderzahl, Einkommen, Körpergröße.

Diskrete Merkmale

- Können nur ganz bestimmte (endlich viele oder abzählbar unendlich viele) abgestufte (Zahlen-)Werte als Merkmalsausprägung haben.
- Diskret sind alle Merkmale, deren Ausprägungen man komplett (auf-)zählen kann, auch wenn (evtl.) keine Obergrenze vorhanden ist (und Zählen unendlich lange dauert).
- Beispiele: Anzahl der Bauteile, Anzahl der Würfe mit einem Würfel.

Stetige Merkmale (bzw. kontinuierliche Merkmale)

- Können in einem Intervall jeden reellen Wert als Ausprägung annehmen (überabzählbar unendlich viele verschiedene mögliche Merkmalsausprägungen innerhalb eines Intervalls).
- Beispiele: Zeit-, Längen- und Gewichtsgrößen.

Nominalskala

- Die Ausprägungen eines nominalskalierten Merkmals können nicht geordnet werden.
→ keine natürliche Reihenfolge.
- Der einzig mögliche Vergleich ist die Prüfung auf Gleichheit zweier Merkmalsausprägungen.
- Beispiel: Obstsorten (Apfel, Banane, Orange, Weintraube, ...)



Ordinal- bzw. Rangskala

- Die Merkmalsausprägungen können gemäß ihrer Größe, Intensität, o.ä. geordnet werden. → Sie haben eine natürliche Reihenfolge
- Eine Interpretation der Rangordnung ist möglich.
- Abstände zwischen den Merkmalsausprägungen können jedoch nicht interpretiert werden.
- Beispiel: Schulnoten (amerikanisches System: A,B,C,D,E,F)



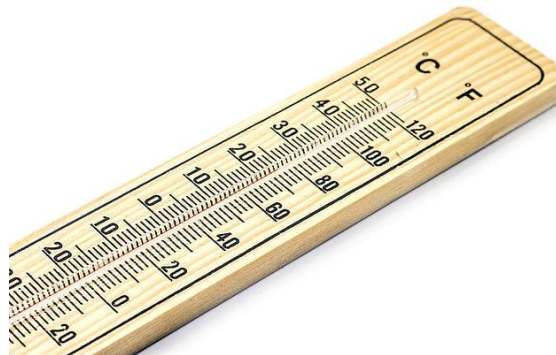


Metrische Skala

- Unter den Merkmalsausprägungen kann eine Rangordnung definiert werden
- Zusätzlich können Abstände zwischen den Merkmalsausprägungen gemessen und interpretiert werden.
- Beispiele: Temperatur, Semesterzahl.
- Metrische Skalen können weiter unterteilt werden in ...
 - Intervallskala
 - Verhältnisskala
 - Absolutskala

Intervallskala

- Es sind nur Differenzenbildungen zwischen den Merkmalsausprägungen zulässig.
- Abstände zwischen Ausprägungen berechenbar und aussagekräftig
- Der Nullpunkt ist willkürlich (nicht natürlich) gewählt
- Daher können nur Abstände verglichen werden.
- Beispiele: Temperatur in C° , Kalender mit Nullpunkt „Christi Geburt“



Verhältnisskala

- Erweiterung der Intervallskala, in der zusätzlich ein natürlicher Nullpunkt existiert.
- Die Bildung von Quotienten ist zulässig
- Verhältnisse sind damit sinnvoll interpretierbar.
- Beispiele: Geschwindigkeit, Längenmaße, Temperatur in Kelvin mit absolutem Nullpunkt





Absolutskala

- Erweiterung der Verhältnisskala, in der zusätzlich auch noch eine natürliche Einheit vorhanden ist.
- Typischerweise, wenn es um eine Anzahl geht
- Beispiele: Semesterzahl, Anzahl Tomaten in einer Kiste



- Zwischen den vorgestellten Skalenarten besteht eine **Rangordnung**, die sich in der Anwendbarkeit statistischer Begriffe und Verfahren widerspiegelt. In aufsteigender Rangfolge sind dies:
 1. Nominalskala,
 2. Ordinal- bzw. Rangskala,
 3. Intervallskala und
 4. Verhältnisskala/Absolutskala.
- Jedes Merkmal kann auch auf einer niedrigeren Skala gemessen werden, dies ist jedoch (üblicherweise) mit einem Informationsverlust verbunden.
- Beispiel: Temperatur
 - Kann auch auf einer Ordinalskala mit den Ausprägungen **kalt**, **normal**, **warm** und **heiß** gemessen werden.
 - Die so gemessenen Temperaturangaben sind aber wesentlich weniger aussagekräftig als Temperaturen auf der Celsius-Skala.



VE 01: Merkmale und Stichproben

1.1 Einstieg: Bundestagswahl 2013

1.2 Grundgesamtheit und Merkmale

1.3 Stichproben und Stichprobenauswahlverfahren

Vollerhebung

Bei einer **Vollerhebung** wird die vollständige Grundgesamtheit hinsichtlich des interessierenden Merkmals untersucht.

- Beispiel: Volkszählung.

Teilerhebung

Es wird eine nach bestimmten Kriterien ausgewählte Teilmenge der Grundgesamtheit hinsichtlich des interessierenden Merkmals untersucht.

- Beispiel: zerstörende Materialprüfung.

Anmerkung: Oftmals sprechen Kosten-, Zeit- oder technische Gründe gegen eine Vollerhebung.

*„Ein Psychiater schrieb einmal, die ganze Menschheit sei verrückt.
Gefragt, wie er denn zu dieser Meinung käme, sagte er:
Sehen Sie sich doch die Leute an, die in meiner Praxis sind ...“ [10,p.99]*

Stichprobe

Die im Rahmen einer Teilerhebung ausgewählte Teilmenge heißt **Stichprobe** und sollte bezüglich des betrachteten Merkmals repräsentativ hinsichtlich der Grundgesamtheit sein.

Notation: Eine Stichprobe x mit n Elementen wird notiert als

$$x = (x_1, \dots, x_n)$$

Die Merkmalsausprägung des i -ten Elements der Stichprobe wird dabei durch x_i dargestellt.



Zufallsstichprobe

Eine **Zufallsstichprobe** ist eine Stichprobe, bei der jedes Element der Grundgesamtheit gleichwahrscheinlich (zufällig) für die Stichprobe ausgewählt wird.

Einfachen Zufallsstichprobe

Für eine **einfachen Zufallsstichprobe** müssen die einzelnen Elemente einer Zufallsstichprobe unabhängig voneinander aus der Grundgesamtheit ausgewählt werden.

Zur Gewinnung einer Stichprobe aus einer Grundgesamtheit gibt es zahlreiche Varianten, u.a.

- Zufallsstichprobe
- geschichtete Zufallsstichprobe
- Klumpenstichprobe
- mehrstufiges Auswahlverfahren
- systematische Auswahl
- Quotenauswahl
- bewusste Auswahl
- u.a.

Aufgabe:

Recherchieren Sie zu jeder dieser Varianten, wie die Ziehung der Stichprobe konkret durchgeführt werden kann.

- [1] Veith Tiemann: *Statistik für Studienanfänger*. Konstanz, München: UVK Verlagsgesellschaft, 2012. (ISBN-13:978-3-8252-3574-1)
- [2] Ludwig Fahrmeier, Rita Künstler, Iris Pigeot, Gerhard Tutz: *Statistik – Der Weg zur Datenanalyse* (7. Auflage). Berlin, Heidelberg: Springer Verlag, 2010.
- [3] Jochen Schwarze: *Grundlagen der Statistik I: Beschreibende Verfahren* (10. Auflage). Herne, Berlin: Verlag Neue Wirtschafts-Briefe, 2005. (ISBN-10:3-482-56430-2)
- [4] Jochen Schwarze: *Grundlagen der Statistik II: Wahrscheinlichkeitsrechnung und induktive Statistik* (8. Auflage). Herne, Berlin: Verlag Neue Wirtschafts-Briefe, 2006. (ISBN-10:3-482-56868-5)
- [5] Ulrich Kockelkorn: *Statistik für Anwender*. Berlin, Heidelberg: Springer Verlag, 2012. (ISBN-13:978-8274-2294-1)
- [6] Thomas Benesch: *Schlüsselkonzepte zur Statistik*. Berlin, Heidelberg: Springer Verlag, 2013. (ISBN-13:978-8274-2771-7)
- [7] Schira, Josef: *Statistische Methoden der VWL und BWL – Theorie und Praxis* (4. Auflage). München: Verlag Pearson Studium, 2012. (ISBN-13:978-3-8689-4117-3)
- [8] Hartung, Joachim; Elpelt, Bärbel; Klösener, Karl-Heinz: *Statistik – Lehr- und Handbuch der angewandten Statistik*. München: Oldenbourg Verlag, 2009. (ISBN-13:978-3-486-59028-9)

- [9] Bleymüller, Josef; Gehlert, Günther; Gülicher, Herbert: *Statistik für Wirtschaftswissenschaftler* (15.Auflage). München: Verlag Franz Vahlen, 2008. (ISBN-13:978-3-8006-3529-0)
- [10] Henze, Norbert: *Stochastik für Einsteiger* (10.Auflage). Wiesbaden: Springer Spektrum, 2013. (ISBN-13:978-3-658-03076-6)
- [11] Krämer, Walter: *So lügt man mit Statistik* (4.Auflage). München: Piper Verlag, 2013. (ISBN-13:978-3-492-26413-6)
- [12] Bosch, Karl: *Elementare Einführung in die angewandte Statistik* (8.Auflage). Wiesbaden: Friedrich Vieweg & Sohn Verlag, 2005. (ISBN-13:978-3-8348-0027-5)