

Statistik I

Prof. Dr. Simone Abendschön

11. Vorlesung am 25.01.24: Einstieg Inferenzstatistik

Plan restliches Semester

- Hintergrund Inferenzstatistik (Warum müssen wir uns damit beschäftigen?)
- Wahrscheinlichkeiten für diskrete und stetige Zufallsvariablen
- Statistische Verteilungen: Rolle der (Standard-) Normalverteilung
- Zentrales Grenzwerttheorem
- Übungsbeispiele

Sitzung 8.2.

Wie soll die letzte Sitzung im Semester inhaltlich gestaltet werden?

Folgende Möglichkeiten:

- **A: Normal weiter im Stoff, Wiederholung Statistik I im SoSe einplanen**
 - B: Wiederholung durch Vorbereitung Ihrer erstellter „Klausurfragen“
 - **C: Wiederholung wichtigste Inhalte durch mich, evtl. Quizformat**
- Ergebnis A und C 50/50
- Wiederholungssession am Do. 25.4. (2. Vorlesungswoche im SoSe 24) durch Tutor

Lernziele

- Sie wissen, warum wir uns mit Wahrscheinlichkeiten beschäftigen
- Sie wissen was eine statistische Verteilung bzw. Wahrscheinlichkeitsverteilung ist
- Sie erweitern Ihre Kenntnisse über die sog. „Normalverteilung“ und wissen wozu sie in der Inferenzstatistik dient
- Sie können Flächenanteile und damit Wahrscheinlichkeiten innerhalb der Normalverteilung berechnen

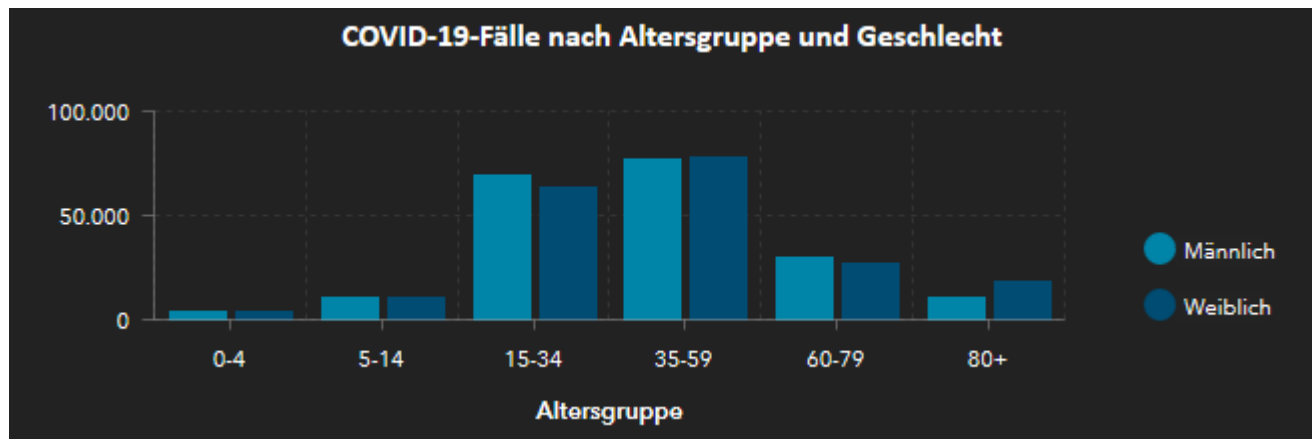
- Forschungsfragen der quantitativen emp. Sozialforschung beziehen sich i.d.R. auf **Grundgesamtheiten** (Ziel: Verallgemeinerung und Prognose)
- Datenerhebung und Datenanalysen werden i.d.R. anhand einer (Zufalls-) **Stichprobe** durchgeführt

Wh. Deskriptive Statistik

- Beschreibende Statistik
- Zentrale Idee: **Informationsverdichtung** und Beschreibung
- Information über eine oder mehrere Variablen wird verdichtet, statt mehrerer 100 oder 1000 Zahlen eine Zahl
- Grafische Darstellungen und Tabellen als Mittel der Veranschaulichung
- Merkmale & Zusammenhänge zwischen Merkmalen werden grundlegend beschrieben

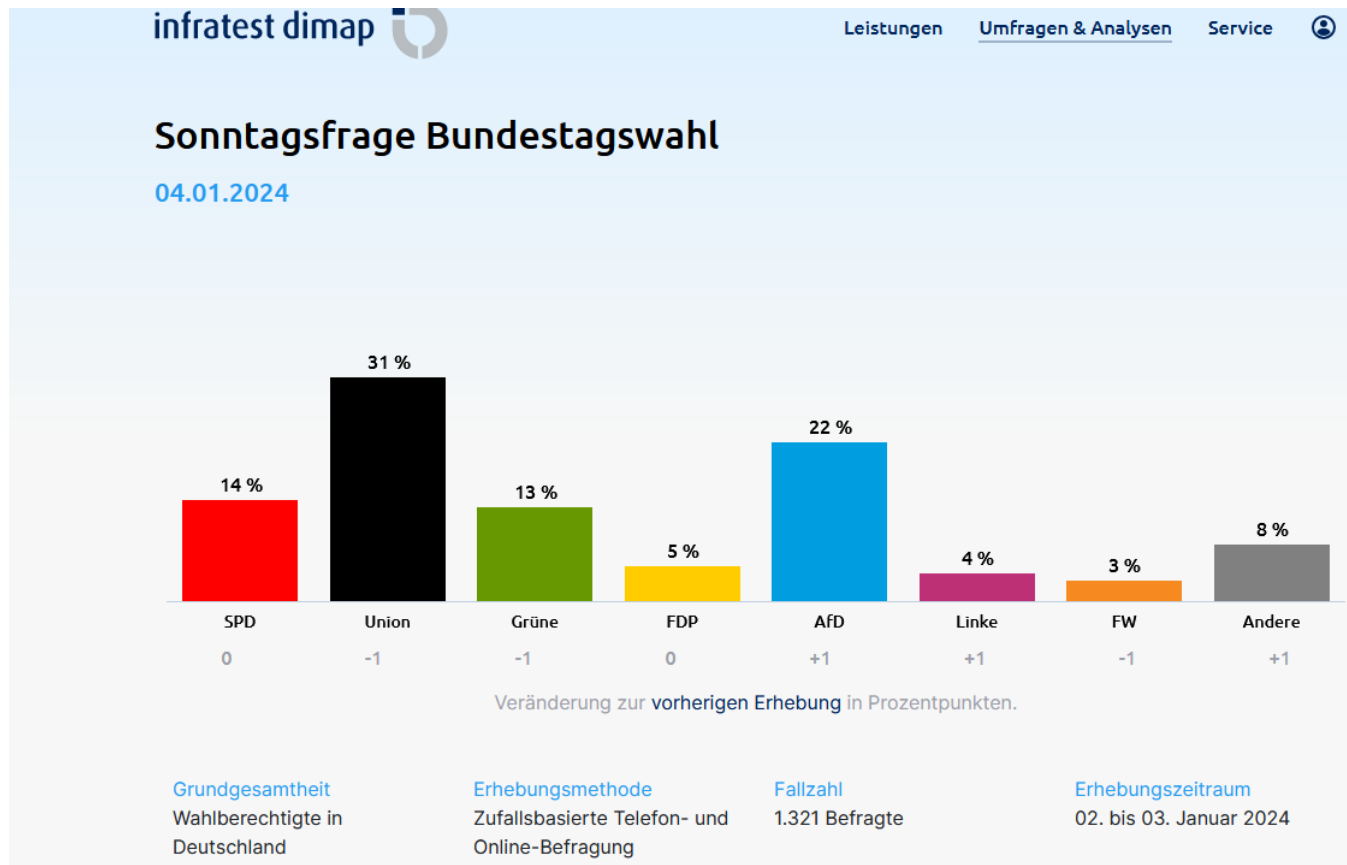
Kennwerte und Parameter („Zahlen“/Maße) für

- Häufigkeiten
- Verteilungen (von Merkmalen bzw. Variablen)
- Streuung und Konzentration von Variablen
- Zusammenhänge (zwischen zwei oder mehreren Variablen)



<https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4>
(Stand 23.10.2020)

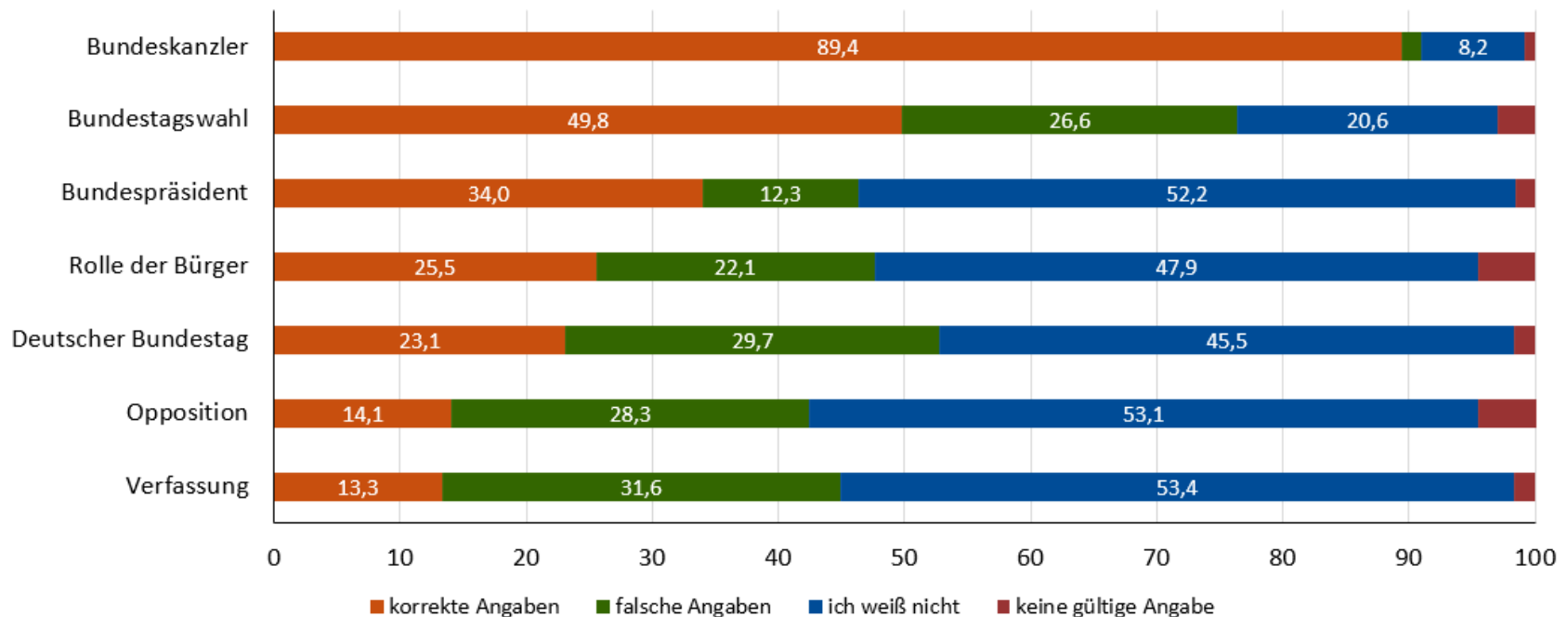
Klassische „Sonntagsfrage“: „Welche Partei würden Sie wählen, wenn am Sonntag Bundestagswahl wäre?“



Quelle: <https://www.infratest-dimap.de/umfragen-analysen/bundesweit/sonntagsfrage/> Stand 24.1.24

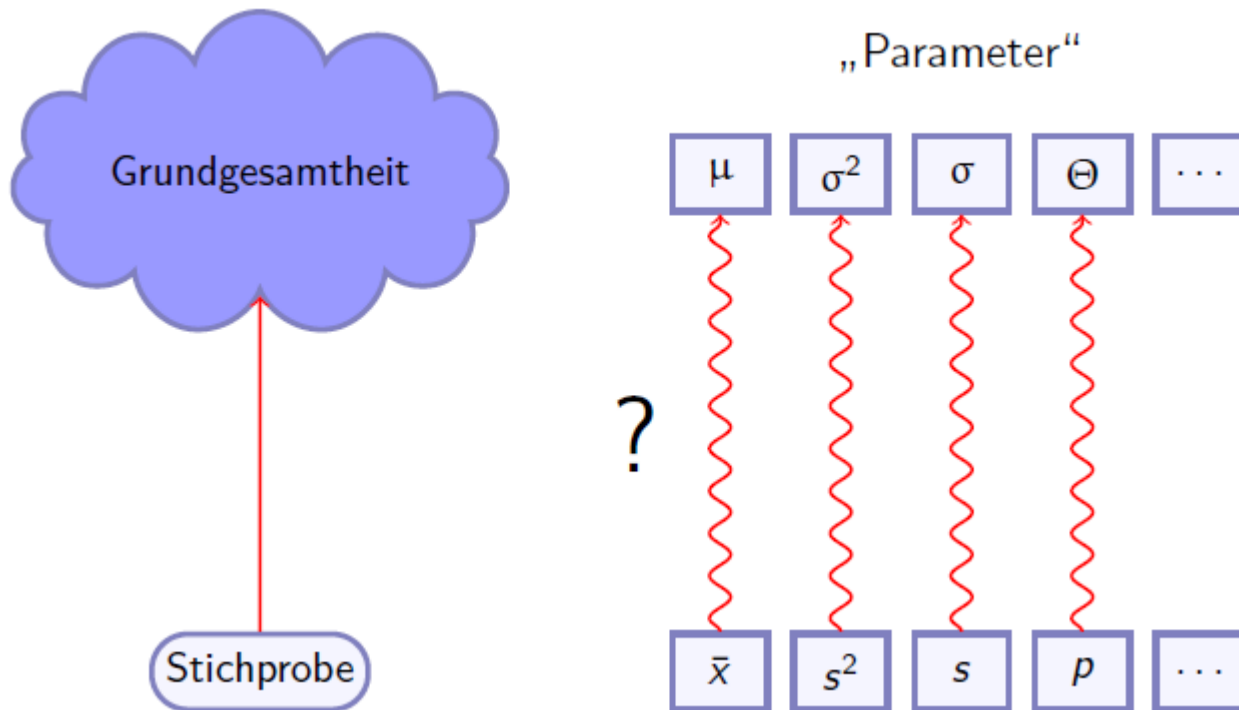
Beispiel aus unserem Vortrag Ringvorlesung

3.2 Erste Ergebnisse: Politisches Wissen



Anmerkungen: n = 1288; Angaben in Prozent; rundungsbedingte Abweichungen von 100 Prozent sind möglich.

Stichprobe und Grundgesamtheit



Population oder Grundgesamtheit

- umfasst die Menge aller möglichen Beobachtungseinheiten, über die man eine Aussage treffen will
- in der Regel sehr groß, kann meistens nicht untersucht werden → Teilerhebung

Stichprobe (sample)

- Teilmenge der (angestrebten) Grundgesamtheit. Sie soll ein möglichst getreues Abbild der Grundgesamtheit sein – das heißt „repräsentativ“

„Repräsentativität“

"Müssen Corona endlich verstehen"

Statistikerin: Wer repräsentative Tests ablehnt, hat die Pandemie nicht verstanden

Teilen

Pocket



Abgerufen auf focus online am 15.4.21

„Repräsentativität“

"Müssen Corona endlich verstehen"

Statistikerin: Wer **repräsentative** Tests ablehnt, hat die Pandemie nicht verstanden

FORDERUNG NACH **REPRÄSENTATIVEN** STUDIEN

Warum Statistiker bei Corona-Tests ein Wörtchen mitreden sollten

von Sabine Hedewig-Mohr

Samstag, 28. März 2020



Abgerufen auf horizont.net am
15.4.20

Repräsentative Umfrage und Stichprobe

Was heißt „repräsentativ“?

- „Meinungsumfragen nennen sich oft ‚repräsentativ‘, tatsächlich aber werden die Befragten meist nach dem Zufallsprinzip ausgewählt.“ (in „Vorwärts“, 10/1994, S. 23 unter der Überschrift „Die Repräsentativitätslüge“)

Repräsentative Umfrage und Stichprobe

- „Meinungsumfragen nennen sich oft ‚repräsentativ‘, tatsächlich aber werden die Befragten meist nach dem Zufallsprinzip ausgewählt.“ (in „Vorwärts“, 10/1994, S. 23)
- Def. „repräsentativ“ quantitative Sozialforschung:
 - Stichprobenkennwerte sind erwartungstreue Schätzer für die Parameter der Grundgesamtheit
 - Voraussetzung: Jedes Element hat die gleiche oder eine angebbare Wahrscheinlichkeit in die Stichprobe zu gelangen (damit Zufallsstichprobe erforderlich)

- Zufallsgesteuerte Auswahlverfahren bzw. Wahrscheinlichkeitsauswahl
- Nicht-zufallsgesteuerte Auswahlverfahren

- Zufallsgesteuerte Auswahlverfahren bzw. Wahrscheinlichkeitsauswahl
 - Auswahlseinheiten werden kontrolliert zufällig aus der Grundgesamtheit entnommen
 - Für alle Einheiten ist die Wahrscheinlichkeit gleich, in die Auswahl aufgenommen zu werden

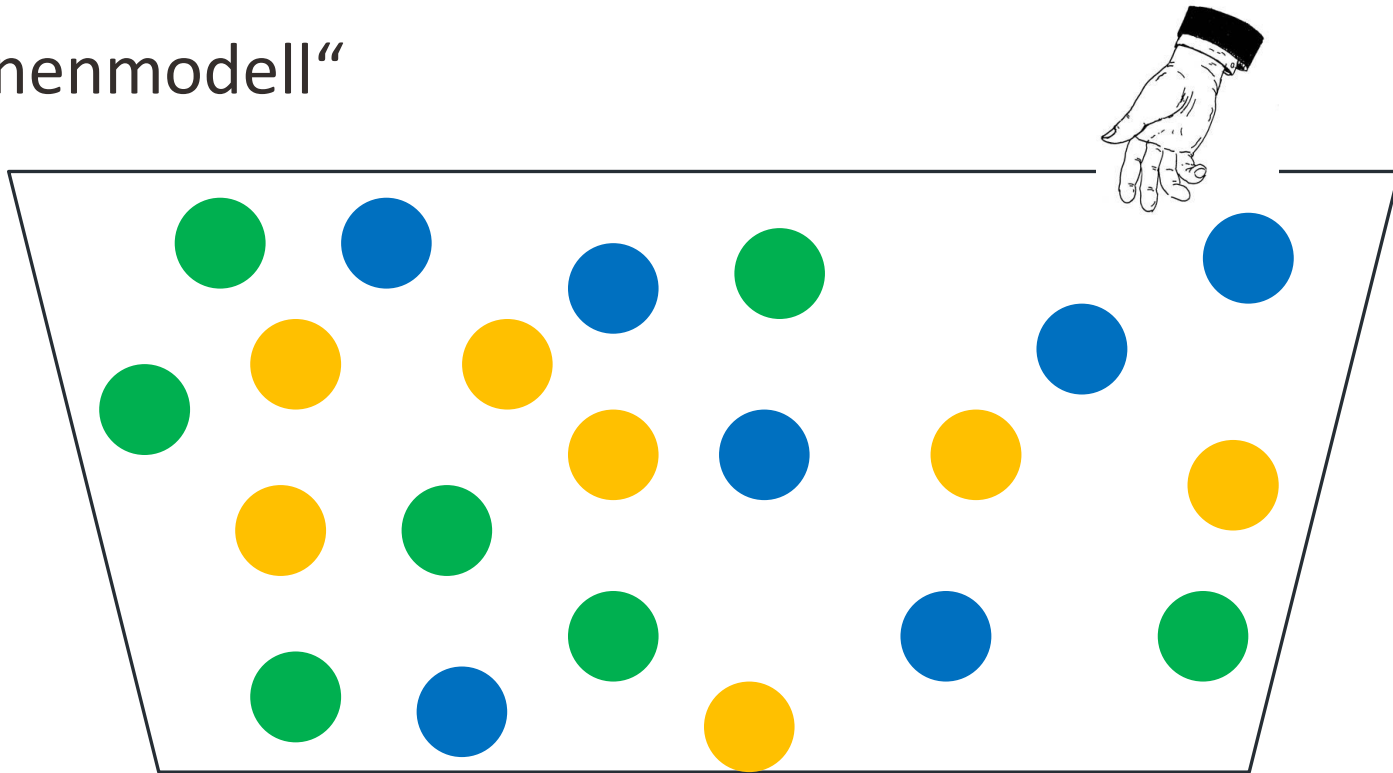
Wh. Zentrale Begriffe - Stichprobenziehung

- Zufallsgesteuerte Auswahlverfahren bzw. Wahrscheinlichkeitsauswahl
 - Einfache Zufallsstichprobe
 - Schichtung
 - Klumpenstichproben

- **Nicht-zufallsgesteuerte Auswahlverfahren**
 - willkürliche Auswahl
 - Quotaverfahren
 - Bewusste Auswahl

Einfache Zufallsstichprobe

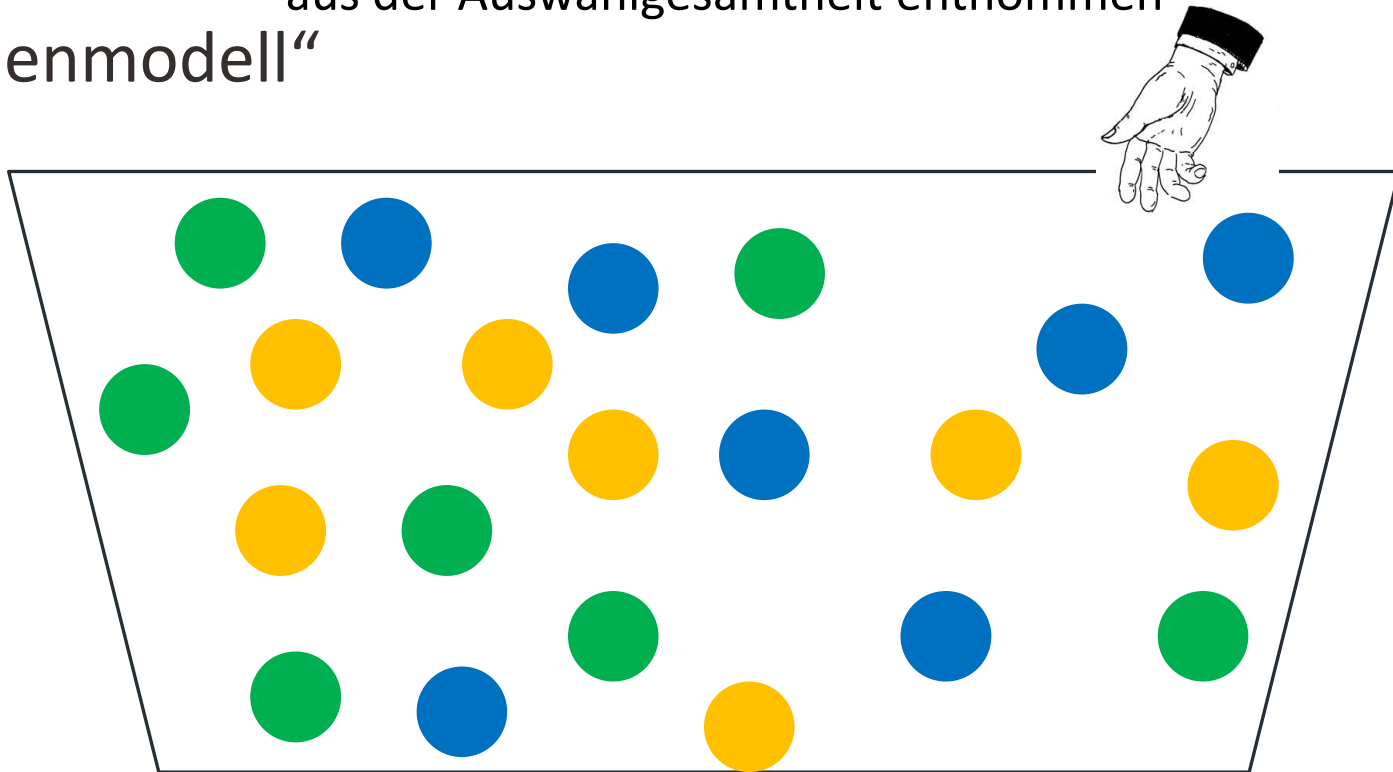
„Urnenmodell“



Einfache Zufallsstichprobe

Untersuchungseinheiten werden zufällig
aus der Auswahlgesamtheit entnommen

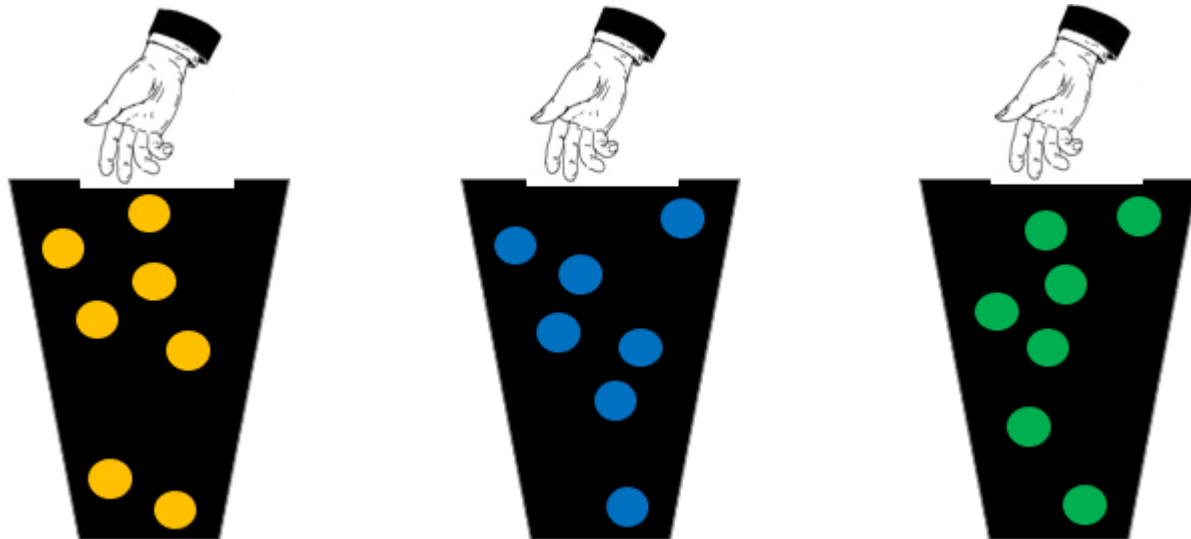
„Urnenmodell“



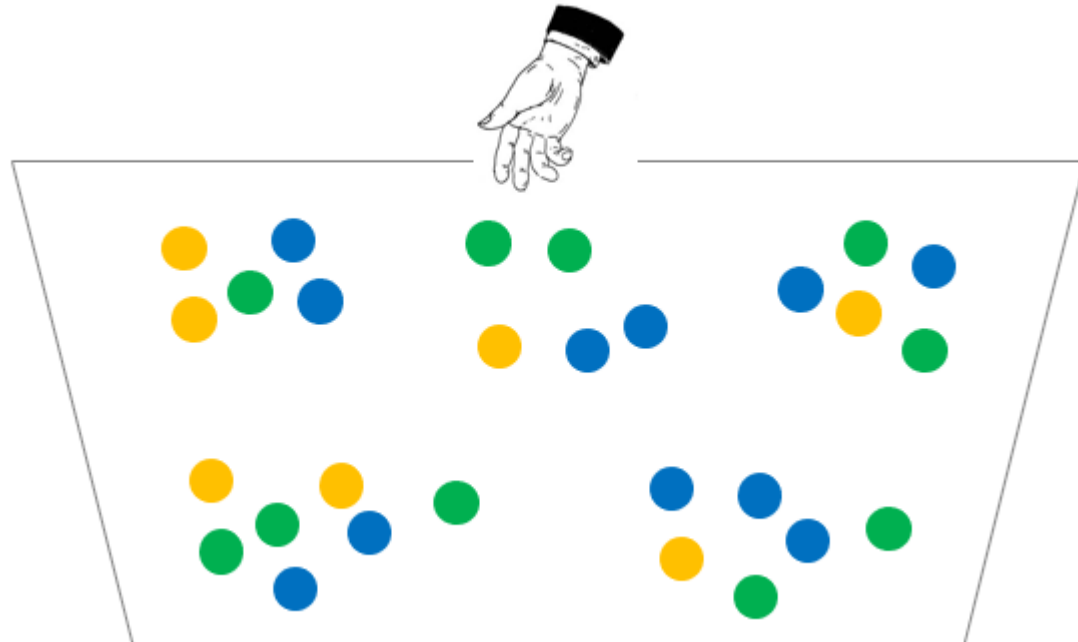
Alle Einheiten haben die gleiche Wahrscheinlichkeit, in die Auswahl aufgenommen zu werden

- In der Praxis wird oft von einer reinen Zufallsauswahl abgewichen: Stufen, Schichten und Klumpen
- Warum?
 - Fehlendes bundesweites Adressenverzeichnis oder Telefonnummernverzeichnis aus denen gezogen werden kann
 - Man möchte Aussagen über eine bestimmte Gruppe treffen oder Vergleiche zwischen sozialen Gruppen vornehmen
 - Forschungspraktische Gründe

- Ziehung einfacher Zufallsstichproben innerhalb von Gruppen der Grundgesamtheit (Schichten), die entsprechend der Forschungsfrage definiert werden



- Ziehung einer (Zufalls-) Stichprobe von Makroeinheiten (Schulen, Organisationen, Haushalten,... = Klumpen), innerhalb derer eine Befragung der Erhebungseinheiten erfolgt



Willkürliche Stichprobe

- Auswahl aufs „Geratewohl“
- Problem: Grundgesamtheit nicht identifizierbar, für quantitative Sozialforschung nicht geeignet

Bewusste Auswahl, „Quotenstichprobe“, Quota-Verfahren

- gezielte Auswahl
- Kenntnisse über Struktur der Grundgesamtheit notwendig
- Quotenvorgabe, z.B. 50 Befragte mit Hauptschulabschluss
- Umstritten: Rückschluss auf Grundgesamtheit

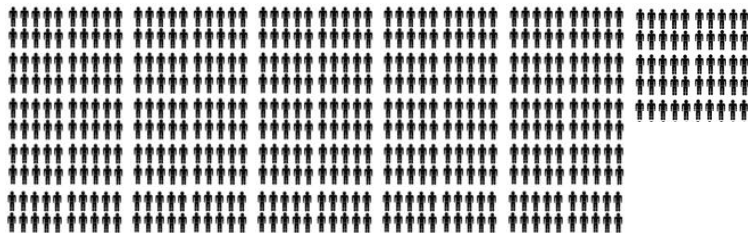
Nicht-zufallsgesteuert

Quota-Verfahren:

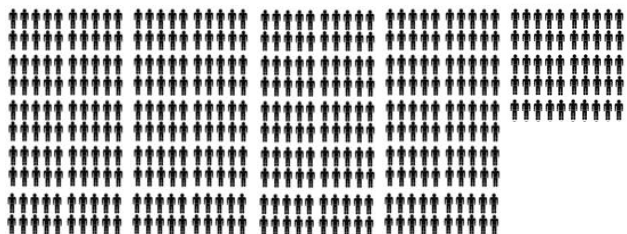
Grundgesamtheit

1000 Personen

550 Männer



450 Frauen



Quote

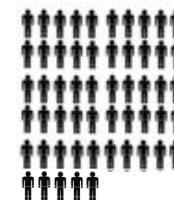
55 %

45
%

Stichprobe

100 Personen

55 Männer

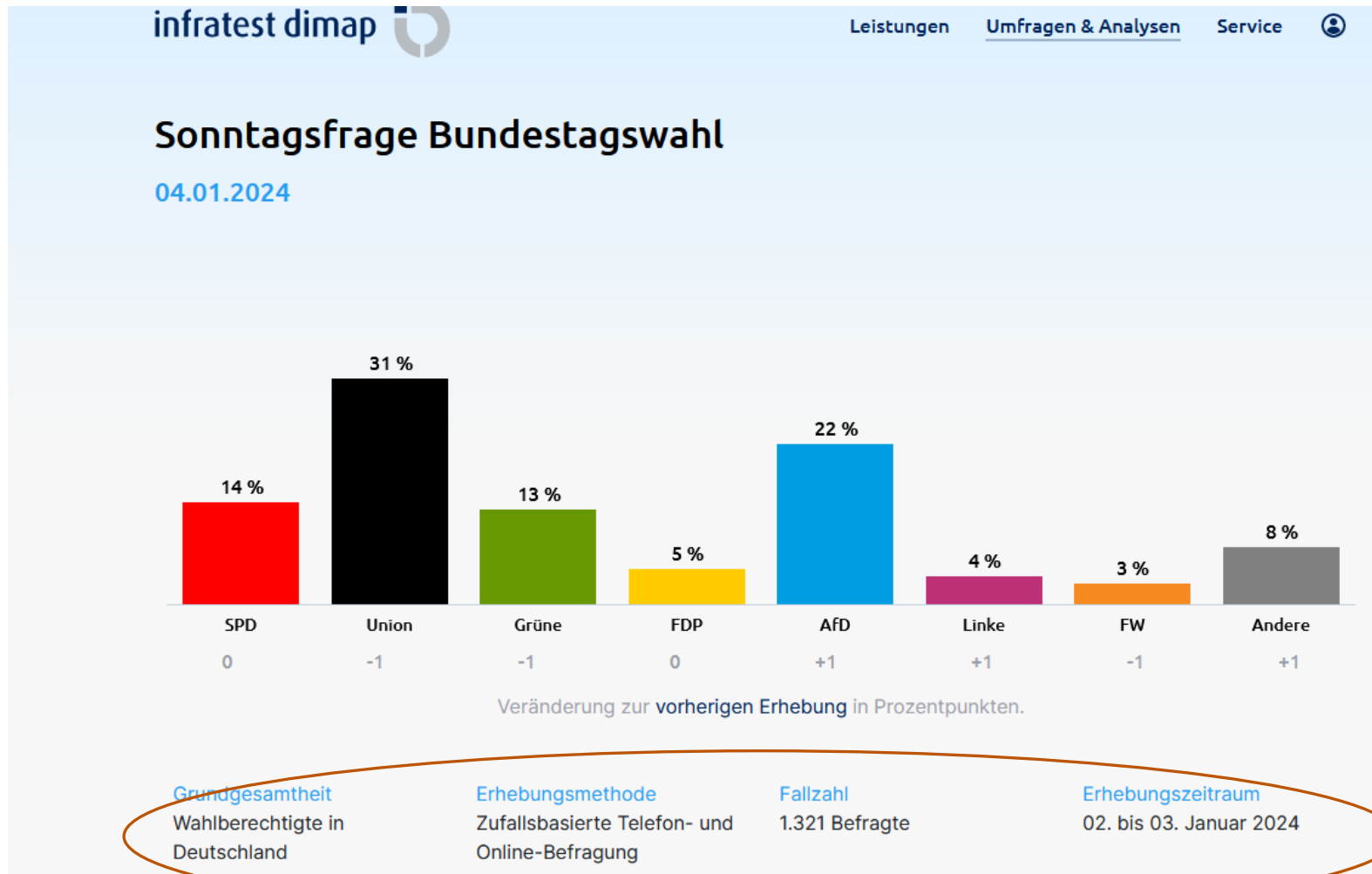


45 Frauen

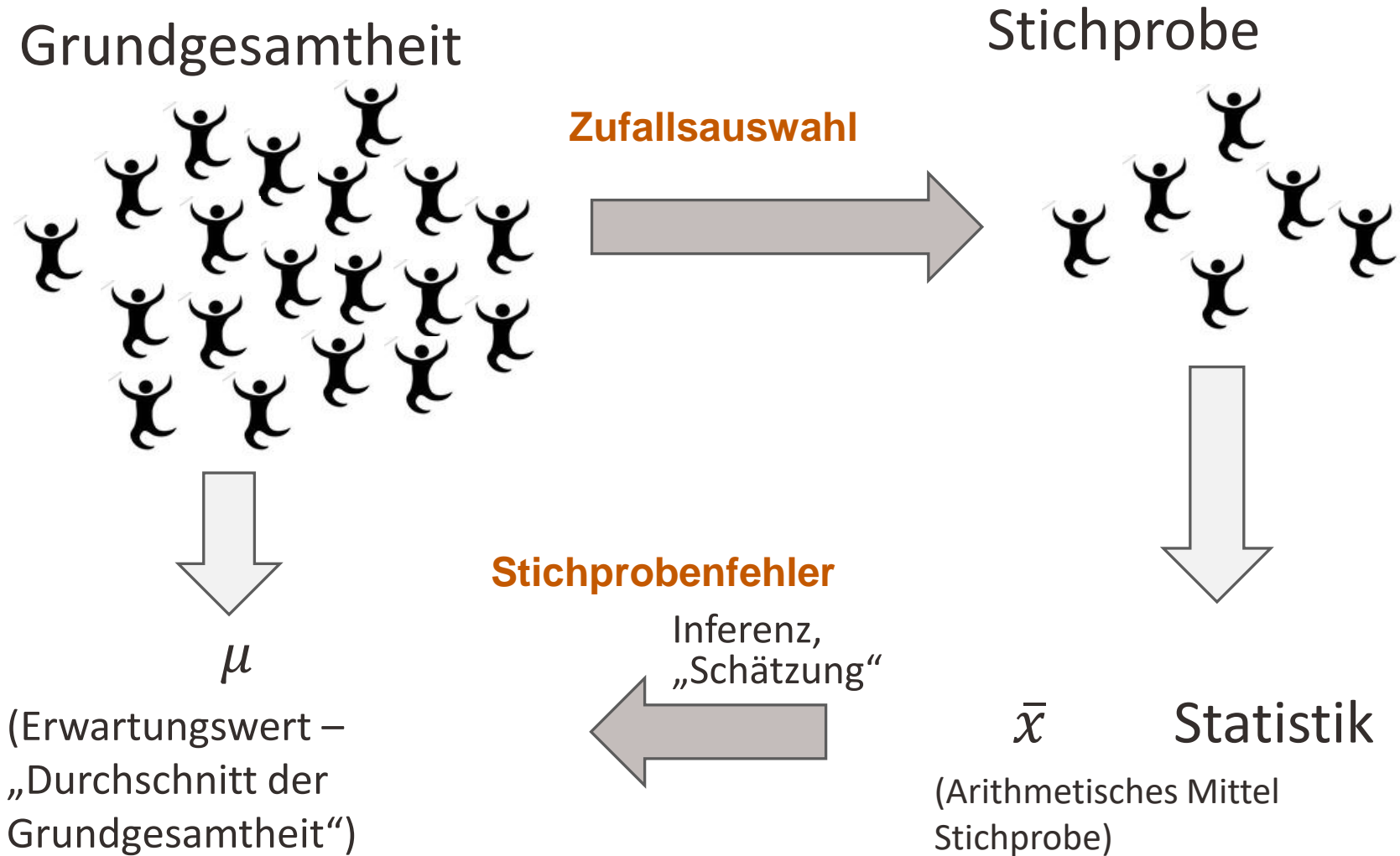


- Stichproben sollen möglichst exaktes Abbild der Grundgesamtheit abgeben, „repräsentativ“ sein
 - Zufallsgesteuerte Auswahlverfahren können das am Besten gewährleisten
 - Nur bei einer Zufallsstichprobe kann innerhalb statistischer Fehlergrenzen ein Befund auf die Grundgesamtheit übertragen werden
 - Nur sie liefern unverzerrte und „erwartungstreue“ Schätzer für die Parameter der Grundgesamtheit

Beispiel Sonntagsfrage infratest dimap



- Primärforscher: GESIS – Leibniz Institut für Sozialwissenschaften
- Datenerhebung: Kantar Public, München
- Erhebungszeitraum: Juni 2021 – August 2021
- Grundgesamtheit: Bevölkerung Deutschlands ab 18 Jahren in Privathaushalten
- Auswahlverfahren: Personenstichprobe (Einwohnermeldeamtsstichprobe), disproportional geschichtete Zufallsauswahl in West-/Ostdeutschland
Erhebungsverfahren: Mixed-Mode Erhebung, Web-Mail Befragung
- (Quelle <https://www.gesis.org/allbus/allbus>)





Notationen und Begriffe

Stichprobe	Wahrscheinlichkeits- verteilungen	Grundgesamtheit
Kennwerte	Parameter	Parameter
Mittelwert \bar{x}	Erwartungswert μ	Mittelwert μ
Standardabweichung s	Standardabweichung σ	Standardabweichung σ
Varianz s^2	Varianz σ^2	Varianz σ^2

Beispiele

- Wie „sicher“ ist es, dass sich ein in einer Stichprobe gefundener Zusammenhang zwischen Bildungsniveau und Wahlentscheidung zugunsten einer Partei auch auf die Grundgesamtheit übertragen lässt?
- Lassen sich Lernerfolge auf ein neu entwickeltes E-Learning-Programm zurückführen oder sind sie dem Zufall geschuldet?
- Schüler*innen mit sozioökonomisch „gutem“ Familienhintergrund haben wahrscheinlich einen höheren Lernerfolg beim „Home-Schooling“ als welche deren Eltern sozioökonomisch schlechter gestellt sind.

- Stochastik

Def. **Wahrscheinlichkeit**:

- Ein Maß für die Chance, dass bei einem **Zufallsexperiment** ein bestimmtes **Ereignis** eintritt.
- Lässt sich in einer (Dezimal-)Zahl zwischen 0 und 1 angeben, wird auch als Bruchzahl, bzw. Prozentwert ausgedrückt (0 bis 100%)

Wahrscheinlichkeit und Zufall

Zufallsexperiment:

- Münz-/Würfelnwurf, Ziehung aus Urne...
- Unter gleichen Bedingungen beliebig oft wiederholbar
- Ausgang unterliegt dem Zufall (kann nicht vorhergesagt werden)

Ereignis:

- Ausgang eines Zufallsexperiments („Kopf oder Zahl“ etc.)
- Es gilt: Ereignisse mit kleiner Wahrscheinlichkeit treten selten auf

Zufallsexperiment A

- Ziehen einer Murmel aus einem Gefäß mit 50 weißen und 50 schwarzen Murmeln
 - 100 Murmeln entsprechen der **Grundgesamtheit**
 - Auszuwählende Murmel entspricht der **Stichprobe**

Zufallsexperiment A

- Ziehen einer Murmel aus einem Gefäß mit 50 weißen und 50 schwarzen Murmeln
 - 100 Murmeln entsprechen der Grundgesamtheit
 - Auszuwählende Murmel entspricht der Stichprobe
- Farbe der gezogenen Murmel kann nicht exakt vorhergesagt werden

Zufallsexperiment A

- Ziehen einer Murmel aus einem Gefäß mit 50 weißen und 50 schwarzen Murmeln
 - 100 Murmeln entsprechen der Grundgesamtheit
 - Auszuwählende Murmel entspricht der Stichprobe
- Farbe der gezogenen Murmel kann nicht exakt vorhergesagt werden
- ABER: die Wahrscheinlichkeit der Auswahl einer schwarzen oder weißen Murmel kann vorhergesagt werden!
 - 50/50 Chance schwarz bzw. weiß (0.5; 0,5, 50%-Wahrscheinlichkeit)

- Auswahl einer Murmel aus einem Gefäß mit 10 weißen und 90 schwarzen Murmeln

- Auswahl einer Murmel aus einem Gefäß mit 10 weißen und 90 schwarzen Murmeln
- Farbe der Murmel kann nicht exakt vorhergesagt werden

- Auswahl einer Murmel aus einem Gefäß mit 10 weißen und 90 schwarzen Murmeln
- Farbe der Murmel kann nicht exakt vorhergesagt werden
- ABER: die **Wahrscheinlichkeit** der Auswahl einer schwarzen oder weißen Murmel kann vorhergesagt werden!
 - Hohe Wahrscheinlichkeit schwarz

Zufallsexperiment C

- Gefäß 1: 50 weiße und 50 schwarze Murmeln,
Gefäß 2: 10 weiße und 90 schwarze Murmeln
- Gegeben sei eine Stichprobe von $n = 4$; alle diese 4 Murmeln seien schwarz

- Gefäß 1: 50 weiße und 50 schwarze Murmeln,
Gefäß 2: 10 weiße und 90 schwarze Murmeln
 - Gegeben sei eine Stichprobe von $n = 4$; alle diese 4 Murmeln seien schwarz
- „Aus welchem Gefäß stammen die Murmeln?“

- Gefäß 1: 50 weiße und 50 schwarze Murmeln, Gefäß 2: 10 weiße und 90 schwarze Murmeln
- Gegeben sei eine Stichprobe von $n = 4$; alle diese 4 Murmeln seien schwarz
- „Aus welchem Gefäß stammen die Murmeln?“
 - Geringe Wahrscheinlichkeit für Gefäß 1, hohe Wahrscheinlichkeit für Gefäß 2 (weil 90 von 100 Murmeln schwarz sind)

Wahrscheinlichkeit

- Berechnung A-priori oder Laplace-Wahrscheinlichkeit für ein Ereignis A
- $$P(A) = \frac{\text{Zahl der für A günstigen Ereignisse}}{\text{Zahl aller möglichen Ereignisse}}$$

(Theoretische) Wahrscheinlichkeit

- Berechnung A-priori oder Laplace-Wahrscheinlichkeit für ein Ereignis A
- $$P(A) = \frac{\text{Zahl der für A günstigen Ereignisse}}{\text{Zahl aller möglichen Ereignisse}}$$

Beispiele:

- Kartenspiel mit 52 Karten, Wahrscheinlichkeit für „Herz König“: $\frac{1}{52}$
- Wahrscheinlichkeit für „Ass“: $P(\text{Ass}) = \frac{4}{52}$

Weitere Beispiele

- „Wie hoch ist die Wahrscheinlichkeit, dass eine zufällig gezogene Karte die Farbe ‚Pik‘ hat?“
 - $P(Pik) = \frac{13}{52} = \frac{1}{4}$
- Brüche können auch in Dezimalschreibweise bzw. in Prozent dargestellt werden:
 - $P(Pik) = \frac{13}{52} = \frac{1}{4} = 0.25 = 25\%$
 - $P(Kopf) = \frac{1}{2} = 0.5 = 50\%$
- Wahrscheinlichkeit von 0 bis 100%

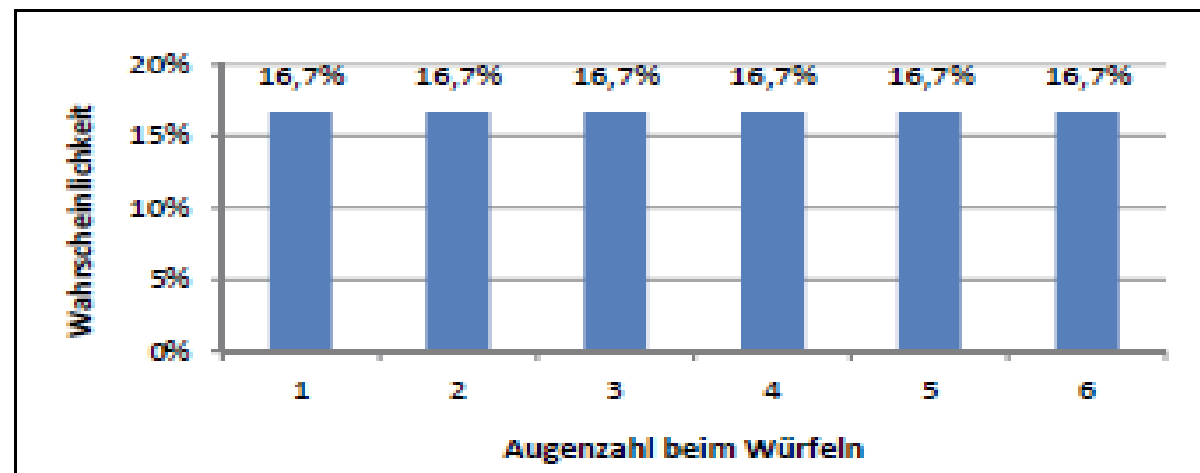
Kleine Übung

- Wie hoch ist die Wahrscheinlichkeit, dass bei einem Wurf mit einem „fairen“ Würfel die Zahl „6“ gewürfelt wird?

Wahrscheinlichkeitsverteilung

- Gibt Wahrscheinlichkeit an, dass ein bestimmtes Ereignis eintritt
- Bei (leicht) abzählbaren Ereignissen in Form von Säulendiagrammen

Abb. 5-2: Wahrscheinlichkeitsverteilung für den Würfelwurf



Quelle: Kuckartz et al. 2013: 121

(Empirische) Wahrscheinlichkeit

- „A posteriori“
- In der wissenschaftlichen Praxis kennt man die Anzahl der günstigen und möglichen Ereignisse oft nicht
- Schätzwert für eine Wahrscheinlichkeit
- Zufallsexperiment wird sehr häufig wiederholt und dabei die relative Häufigkeit von Ereignisausgängen ermittelt (heutzutage mit Computersimulation)

Wahrscheinlichkeit und relative Häufigkeiten

- Zusammenhang zwischen Wahrscheinlichkeit und relativen Häufigkeit
- Wahrscheinlichkeit für ein Ereignis entspricht genau seinem relativen Anteil am Ereignisraum

Wahrscheinlichkeit und relative Häufigkeiten

- Beispiel:
 - 100 Teilnehmer einer politikwissenschaftlichen Univeranstaltung, 70 aus BASS-Studiengang, 30 aus Master.

Wahrscheinlichkeit und relative Häufigkeiten

- Beispiel:
 - 100 Teilnehmer einer politikwissenschaftlichen Univeranstaltung, 70 aus BASS-Studiengang, 30 aus Master.
 - Sie wählen zufällig eine Person für eine Befragung aus. Wie hoch ist die Wahrscheinlichkeit, dass diese Person aus dem BA kommt?

Wahrscheinlichkeit und relative Häufigkeiten

- Beispiel:
 - 100 Teilnehmer einer politikwissenschaftlichen Univeranstaltung, 70 aus BASS-Studiengang, 30 aus Master.
 - Sie wählen zufällig eine Person für eine Befragung aus. Wie hoch ist die Wahrscheinlichkeit, dass diese Person aus dem BA kommt?
 - Relative Häufigkeit für BA: $70/100=7/10=0,7 \rightarrow 70\%$ Wahrscheinlichkeit

Wahrscheinlichkeit und relative Häufigkeiten

- Anwendungsbeispiel: Wahrscheinlichkeit und relative Häufigkeit

Tabelle: Wirksamkeit von Comirnaty

Wie viele Personen erkrankten an Covid-19?	Placebo	Comirnaty	Prozentuale Verringerung des Risikos
Alle Teilnehmenden	93 von 10.000	5 von 10.000	ca. 95 %
16 bis 55 Jahre	115 von 10.000	5 von 10.000	ca. 96 %
Über 55 Jahre	64 von 10.000	4 von 10.000	ca. 94 %

Quelle: <https://www.gesundheitsinformation.de/der-impfstoff-comirnaty-bnt162b2-biontech-pfizer-zur-impfung-gegen-corona.html> (nicht mehr online)