

Statistik - Grundlagen

Es gibt drei Arten von Lügen:

Lügen, infame Lügen und Statistiken.

Benjamin Disraeli, britischer Staatsmann und Schriftsteller

Es gibt drei Arten von Lügen: Lügen, infame Lügen und Statistiken.

Leonard Henry Courtney, britischer Politiker und Schriftsteller

Das Zitat wurde fälschlicherweise von Mark Twain Disraeli zugeordnet, ist aber tatsächlich nachweislich von Courtney.

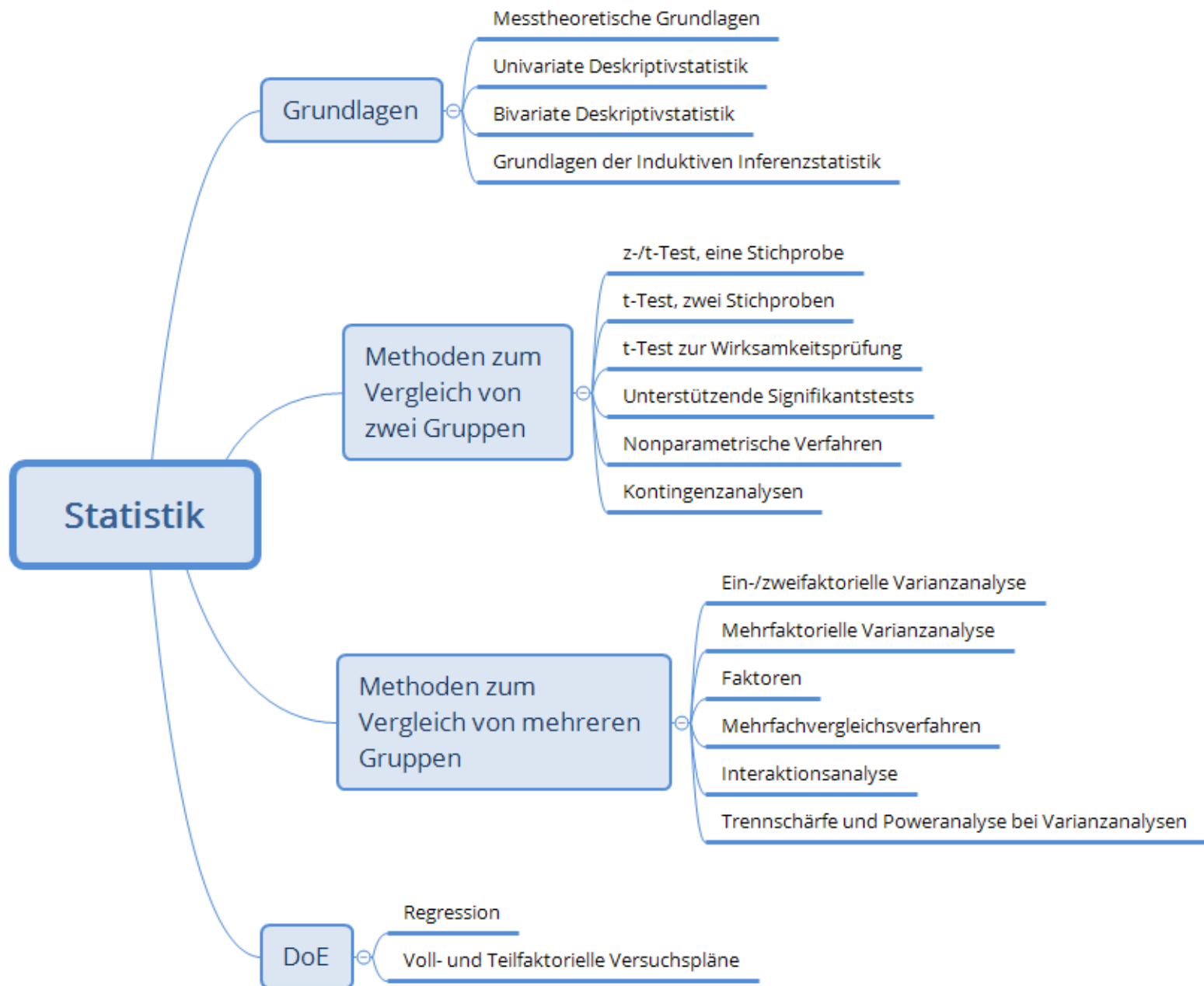
Man kann sich nie ganz sicher sein...

Wenn vor Jahren schon die Zahl der Brücken veröffentlicht wurde, die in den nächsten Jahren einstürzen werden, und diese Brücken dennoch einstürzen, ist damit nichts gegen die Statistik gesagt, sondern einiges über die bedauerliche Tatsache, dass die richtigen Zahlen nie von den richtigen Leuten zur rechten Zeit gelesen werden.

Dieter Hildebrandt, deutscher Kabarettist

Für mich das Informationsmittel der Mündigen.
Wer mit ihr umgehen kann, kann weniger
leicht manipuliert werden.
Der Satz "Mit Statistik kann man alles
beweisen" gilt nur für die Bequemen, die keine
Lust haben, genau hinzusehen

Elisabeth Noelle-Neumann, dt. Marktforscherin



Einige Grundbegriffe

Sta·tis·tik /Statístik/

Substantiv, feminin [die]

1. [ohne Plural]

Wissenschaft von der zahlenmäßigen Erfassung,
Untersuchung und Auswertung von Massenerscheinungen

2. schriftliche Zusammenstellung der Ergebnisse von
Massenuntersuchungen (meist in Form von Tabellen oder
Grafiken)

Definition lt. Oxford Languages / Google

Einige Grundbegriffe

Statistik „ist die Lehre von Methoden zum Umgang mit quantitativen Informationen“ (Daten).

Sie ist eine Möglichkeit, „eine systematische Verbindung zwischen Erfahrung und Theorie herzustellen“.

Unter Statistik versteht man die Zusammenfassung bestimmter Methoden, um empirische Daten zu analysieren.

(Quelle: Wikipedia)

Einige Grundbegriffe

Beschreibende Statistik (Deskriptive Statistik)

Daten werden in geeigneter Weise beschrieben, aufbereitet und zusammengefasst

Verdichtung von quantitativen Daten zu Tabellen, grafischen Darstellungen und Kennzahlen

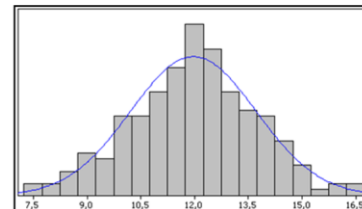
1	15,1152	10,4292	8,0701	10,2995	12,7008
2	12,3493	12,4727	11,4234	11,0104	8,6983
3	13,2927	9,6465	16,7010	12,9357	12,6594
4	14,1500	12,5281	10,2236	13,9782	9,2103
5	13,6990	11,0482	9,1129	14,3687	8,4305
6	9,9523	12,4958	12,6020	12,7010	10,5677
7	12,9891	13,0010	11,0266	12,2196	15,6553
8	12,3055	11,5496	10,1620	7,7583	11,5752
9	13,9653	13,6638	11,1491	9,3201	14,3487

Daten

Kennwerte

Mittelwert	11,972
StdAbw	1,755
Varianz	3,079

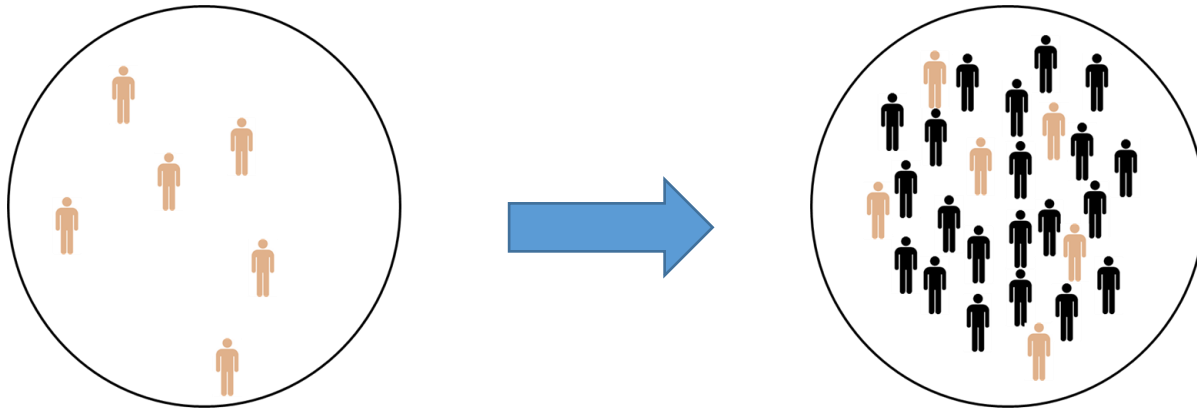
Grafiken



Einige Grundbegriffe

Schließende Statistik (Induktive Statistik)

In der schließenden Statistik leitet man aus den Daten von Stichproben Eigenschaften einer Grundgesamtheit ab



Stichprobenparameter

\bar{x} , s , p , s^2 , n

Populationsparameter

μ , σ , P , σ^2 , N

Einige Grundbegriffe

Grundgesamtheit/Population: Menge aller statistischen Einheiten (Merkmalsträger) mit übereinstimmenden Identifikationskriterien

Alle aktuellen Teilnehmer an
alfatraining-Kursen

Stichprobe: Teilmenge der Population, die tatsächlich untersucht wird

Der Statistik-Kurs

Statistische Einheit: Das einzelne, untersuchte Objekt

Lieschen Müller

Merkmal: Größe, die untersucht werden soll

Körpergröße

Ausprägung: Wert der Größe für ein Individuum

1,63 m

Einige Grundbegriffe

Übung Grundbegriffe

Eine Wohnungsbaugesellschaft will sich zukunftsorientiert aufstellen und möchte deshalb ihre Produktpalette alterskonform umgestalten. Deshalb möchte sie ältere Kunden (60 aufwärts) zu ihren Vorstellungen befragen.

Aus der Kundendatei werden dazu per Zufall 100 Personen über 60 Jahren ausgewählt und kontaktiert.

Erfragt wird unter anderem die Notwendigkeit barrierefreier Zugänge, die einfach mit Ja oder Nein beantwortet werden kann.

Einige Grundbegriffe

Übung Grundbegriffe

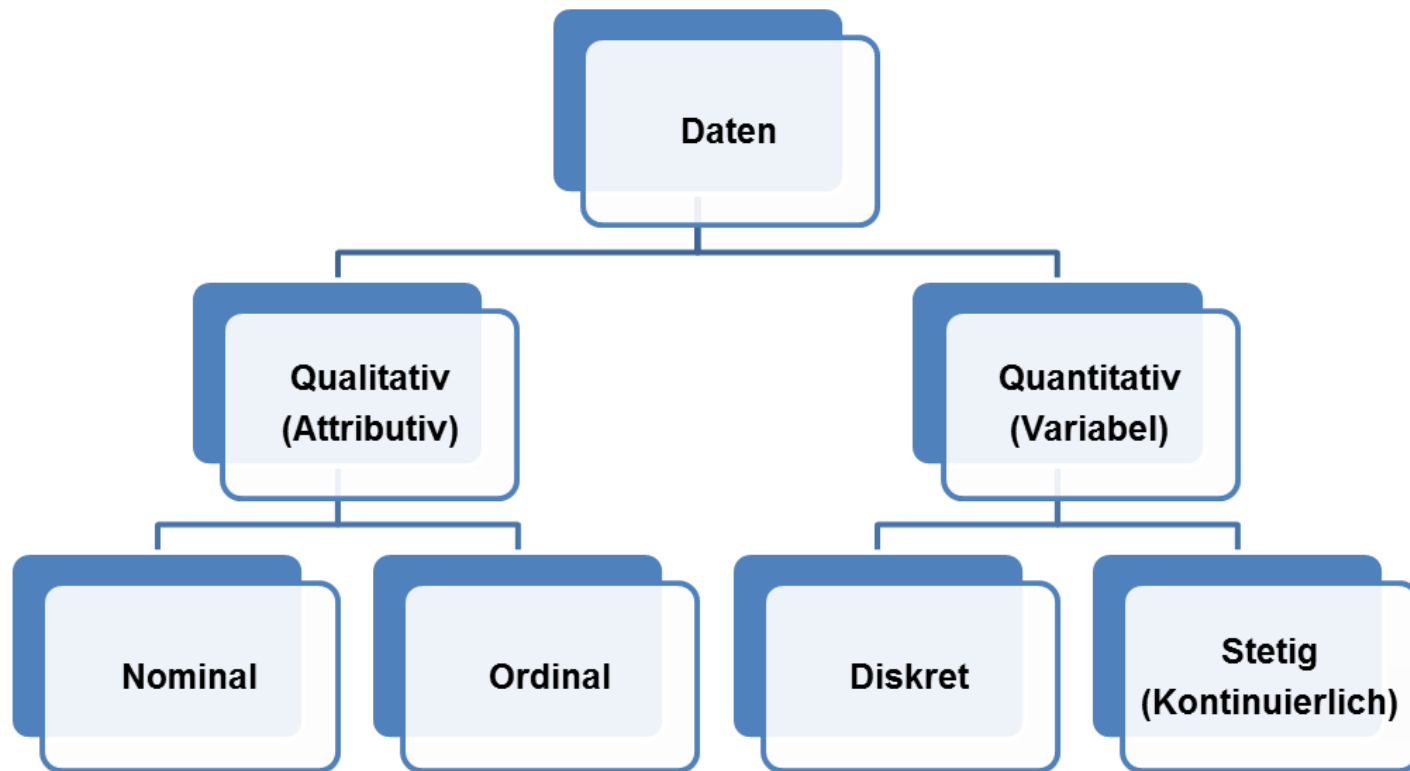
- Grundgesamtheit:
- Teilgesamtheit:
- Stichprobe:
- Statistische Einheit:
- Merkmal:
- Ausprägung:

Einige Grundbegriffe

Übung Grundbegriffe

- Grundgesamtheit: Alle Mieterinnen und Mieter der WBG
- Teilgesamtheit: Alle Mieterinnen und Mieter älter 60 Jahre
- Stichprobe: 100 zufällig ausgewählte Mieter alter 60 Jahre
- Statistische Einheit: Eine einzelne Mieterin/Mieter
- Merkmal: Einschätzung der Notwendigkeit der Barrierefreiheit
- Ausprägung: Notwendig / Nicht notwendig

Klassifizierung von Daten



Zunehmende Datenqualität →

Klassifizierung von Daten

Qualitative / attributive Daten

- Im physikalischen Sinne nicht messbar
- Geben Merkmalseigenschaften wieder
- Können in Kategorien zusammengefasst werden
- Zählbar
- Beispiele: Frau / Mann / Kind; gut / mittel / schlecht; groß / klein; hell / dunkel; rot / gelb / blau; Schulnoten

Klassifizierung von Daten

Nominal

- Qualitative / attributive Merkmale ohne natürliche Rangfolge
- Beispiel: Autofarben (grün, blau, rot,...);
Unternehmensstandorte (Köln, Hamburg, Nürnberg,...)

Ordinal

- Qualitative / attributive Merkmale mit natürlicher Rangfolge, ohne definierten Abstand
- Beispiele: i.o / n.i.o; Schulnoten; groß / klein

Klassifizierung von Daten

- Qualitative / attributive Daten haben aus statistischer Sicht einen geringen Informationsgehalt als quantitative / variable Daten
- Eine geeignete Aussagesicherheit ist nur durch große Stichproben zu erreichen
- Nur wenige statistische Verfahren stehen zur Verfügung
- Typische Verteilungsfunktionen :Poisson, Binomial, Hypergeometrisch

Klassifizierung von Daten

Quantitative Daten

- Im physikalischen Sinne messbar bzw. abzählbar in unendlich vielen Ausprägungen
- **Diskrete Daten**
 - Klassifizierbar, basieren auf Zählung
 - Beispiele: Fehleranzahl, Anzahl Kundenreklamationen
- **Kontinuierliche Daten**
 - Können auf kontinuierlichen Skalen gemessen werden
 - Können jeden möglichen Zwischenwert annehmen
 - Beispiele: Länge, Zeit, Gewicht

Skalenniveaus

Um die Ausprägung eines Merkmals messen zu können, muss man eine Skala festlegen, die alle möglichen Ausprägungen eines Merkmals umfasst

Nominalskala

- Besteht aus Namen, Kategorien oder Kennzeichnungen
- Werte unterliegen keiner Rangfolge
- Arithmetische Operationen sind nur eingeschränkt zulässig
- Beispiel: $m / w / d$

Skalenniveaus

Ordinalskala

- Werte unterscheiden sich in Intensität und können danach geordnet werden
- Abstände sind nicht interpretierbar
- Beispiel: *Schulnoten*

Skalenniveaus

Intervallskala (Metrische Skala)

- Werte unterscheiden sich in Intensität und können danach geordnet werden
- Abstände sind interpretierbar
- Nullpunkt und Abstand sind willkürlich
- Beispiel: *Temperatur in Celsius*

Skalenniveaus

Ratioskala / Verhältnisskala (Metrische Skala)

- Werte unterscheiden sich in Intensität und können danach geordnet werden
- Abstände sind interpretierbar
- Nullpunkt und Abstände sind definiert
- Verhältnisse können gebildet werden
- Beispiel: *Durchmesser, Temperatur in Kelvin, Fehleranzahl*

Skalenniveaus

Übung: Skalenniveaus und Variablentyp

	Skalen- niveau	Daten- klasse
Höhe eines Gartenzauns		
Farbe eines PKW		
Umgebungstemperatur		
Schulnoten		
Preis eines Rasenmähers		
Telefonnummer		

Skalenniveaus

Übung: Skalenniveaus und Variablentyp

	Skalen- niveau	
Höhe eines Gartenzauns	metrisch	1,20 m
Farbe eines PKW	nominal	RAL3003 ruby red
Umgebungstemperatur	metrisch	4°C / 40°F
Schulnoten	ordinal	gut / 2
Preis eines Rasenmähers	metrisch	349,99 €
Telefonnummer	nominal/ ordinal	0123/4567 ?

Datenerhebung

Primärstatistik

Auswertung von Daten, die explizit für diesen Zweck gesammelt wurden, z.B. Volkszählung

Sekundärstatistik

Auswertung von vorhandenen Daten, die nicht für den vorliegenden Zweck gesammelt wurden, z.B. Qualitätsdaten

Tertiärstatistik

Auswertung von Daten, die nur komprimiert vorliegen, z.B. Arbeitslosenstatistik

Datenerhebung

Methodik – Wie gewinne ich Daten

Befragung, Beobachtung, Erfassung, Experiment

Ablauf – Spielt die Zeit eine Rolle

- Querschnitt: Erfassung zu einem Zeitpunkt
- Längsschnitt: Wiederkehrende Erfassung

Umfang – Wie viele Daten werden erhoben

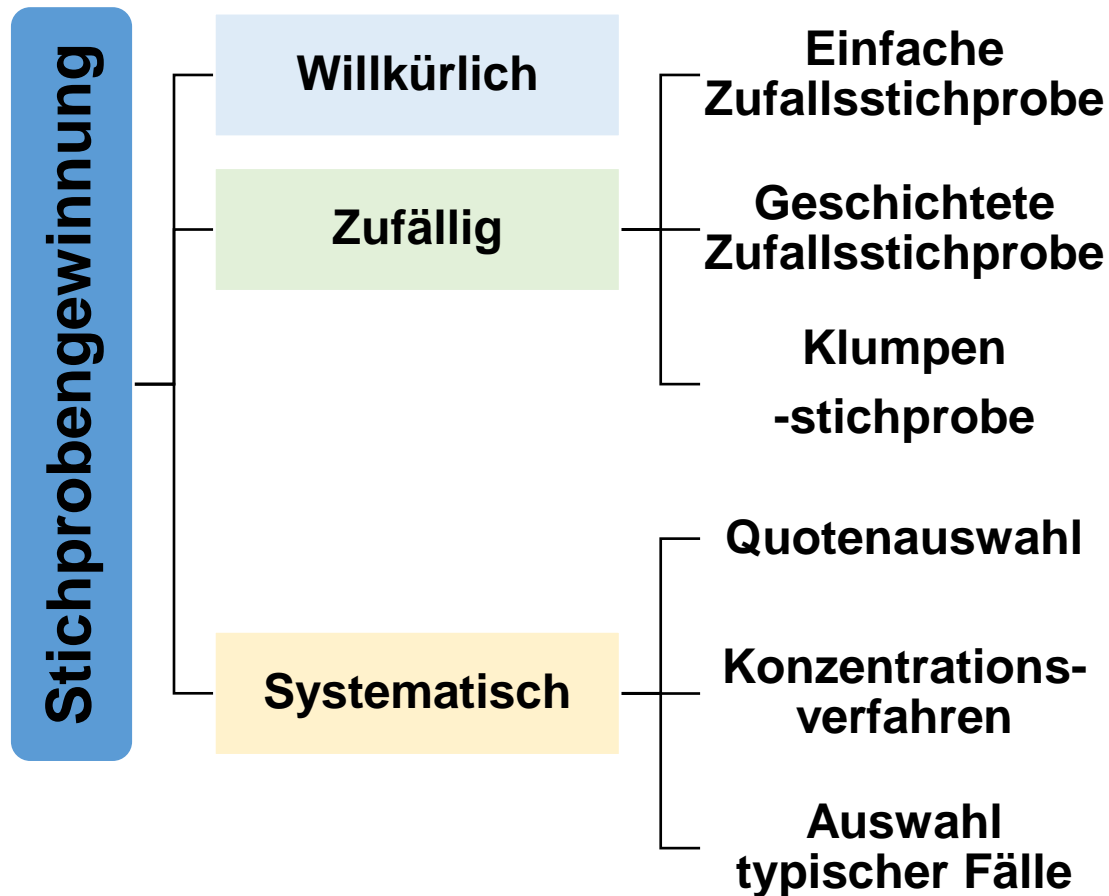
- Vollerhebung: Untersuchung der vollständigen Population
- Teilerhebung: Untersuchung von Stichproben

Stichproben

Stichprobe

- Teilmenge einer Grundgesamtheit, die unter bestimmten Aspekten ausgewählt wurde
- Aus den Eigenschaften der Stichprobe können Rückschlüsse auf die Grundgesamtheit gezogen werden
- Stichproben (Teilerhebung) werden genutzt, wenn die Vollerhebung nicht praktikabel bzw. wirtschaftlich vertretbar ist
- Beispiel: Wählerbefragung am Wahllokal

Stichprobengewinnung



Stichprobengewinnung

Willkürliche Stichprobe (convenience sampling)

- Auswahl der Stichprobe folgt keinen festgelegten Regeln
- Nicht mit der Zufallsstichprobe verwechseln
- Ergebnisse i.a. nicht auf die Grundgesamtheit übertragbar
- Oft für Vorstudien eingesetzt
- Einfaches Verfahren
- Beispiel: Straßenumfrage

Stichprobengewinnung

Zufällige Stichprobe

Einfache Zufallsstichprobe (Simple Random Sampling)

- Jedes Element der Grundgesamtheit hat die gleiche Chance ausgewählt zu werden
- Jedem Element wird eine Zahl zugeordnet, anschließend erfolgt die Auswahl über Zufallszahlen
- Beispiel: Daten des Einwohnermeldeamtes

Stichprobengewinnung

Zufällige Stichprobe

Geschichtete Zufallsstichprobe

- Unterteilung der Grundgesamtheit in kleinere, in sich homogene Gruppen (Schichten)
- Innerhalb der Schichten werden einfache Zufallsstichproben erfasst
- Verfahren zur Varianzreduktion
- Beispiel: Medikamentenwirksamkeit bei Rauchern / Nichtrauchern

Stichprobengewinnung

Zufällige Stichprobe

Klumpenstichprobe

- Unterteilung der Grundgesamtheit in kleinere Gruppen (Klumpen)
- Klumpen sind oftmals regional definiert
- Vollerhebung / Teilerhebung erfolgt nur in einzelnen Klumpen
- Gefahr nicht-repräsentativer Klumpen
- Beispiel: Verbraucherverhalten in einer einzelnen Gemeinde wird als Abbild der Grundgesamtheit angenommen

Stichprobengewinnung

Systematische Stichprobengewinnung

Quotenauswahl

- Gezielte Zusammensetzung einer Stichprobe nach quotierten Merkmalen der Grundgesamtheit, z.B. Geschlecht
- Im engeren Sinn keine Zufallsstichproben
- Beispiel: Interviewer hat Personen mit einem bestimmten Alter, Geschlecht oder Beruf zu befragen

Stichprobengewinnung

Systematische Stichprobengewinnung

Konzentrationsverfahren

- Erhebung erfolgt nur bei wesentlichen Einheiten
- Cut-off von unbedeutenden Einheiten
- Einengung der Grundgesamtheit ohne negativen Einfluss auf die Ergebnisse zu nehmen
- Beispiel: Decken wenige Großunternehmen einen Bedarf fast vollständig, lässt man Kleinunternehmen bei einer Befragung außen vor

Stichprobengewinnung

Systematische Stichprobengewinnung

Auswahl typischer Fälle

- Eingrenzung einer Erhebung auf typische Vertreter der Grundgesamtheit
- „Otto Normalverbraucher“
- Beispiel: Befragung von „typischen Kunden“

Stichprobengewinnung

	Willkürlich	Zufällig	Systematisch
Planungs- aufwand	Gering	Hoch	Mittel
Kosten	Gering	Hoch	Mittel
Theoretische Fundierung	Nein	Ja	Nein
Fehler- rechnung	Nein	Möglich	Nein
Qualität	Gering	Hoch	Mittel

Stichproben

Repräsentative Stichprobe

Eine Stichprobe wird so gewählt, dass aus ihr allgemeingültige Rückschlüsse auf die Grundgesamtheit /Population gezogen werden können

Somit muss die Stichprobe die Grundgesamtheit in wesentlichen Merkmalen widerspiegeln