



Skript Statistik/Datenanalyse

Übersicht

- A Einführung
- B Univariate deskriptive Statistik
- C Bi-/Multivariate deskriptive Statistik
- D Schließende Statistik

Inhalt

A1 Definitionen Statistik/Datenanalyse	2
A2 Statistik im Alltag	2
A3 Vollerhebung und Stichprobe.....	3
A4 Merkmale und Merkmalsträger	3
A5 Skalen	4
A6 Datenmatrix	5
B1 Absolute und relative Häufigkeit	7
B2 Modus, Quartile, Median, arithmetisches Mittel	8
B3 Spannweite, Quartilsabstand, Varianz, Standardabweichung	10
C1 Unabhängige und abhängige Variablen, Faktoren.....	12
C2 Kontingenztafel bzw. Kreuztafel	12
C3 Zusammenhangsmaße	14
C4 Regressionsanalyse und Pfadmodell.....	17
C5 Faktorenanalyse, Skalenhomogenität, Reliabilität.....	20
D1 Mittelwert (Schätzung).....	22
D2 Zusammenhänge (Tests).....	24

A Einführung

A1 Definitionen Statistik/Datenanalyse

Statistik:

Sammeln, Analyse, Interpretation oder Präsentation von Daten¹
(in diesem Skript: Fokus auf die Analyse von Daten, also **Datenanalyse**)

Daten:

Ergebnisse von Messvorgängen²

Messen:

Eigenschaften von Objekten werden nach bestimmten Regeln in Zahlen ausgedrückt³

A2 Statistik im Alltag

Statistik begegnet uns im Alltag vor allem in den Medien. Wenn Ergebnisse aus Studien („Statistiken“) in den Medien wiedergegeben werden, fehlen meist wichtige Angaben, die es erlauben würden, die Ergebnisse sachgerecht zu interpretieren.



Grafik: Beispiel für die Darstellung von Daten in den Medien⁴

Es fehlt oft die Angabe, welche Datengrundlage die Studie hatte (z. B. wie viele Menschen wie befragt wurden), wie die Daten erhoben wurden (z. B. Befragung oder Sekundäranalyse) oder worauf sich Prozentangaben beziehen (z. B. Anteil der Befragten oder Anteil an Gesamtnutzung).

¹ <http://de.wikipedia.org/wiki/Statistik>

² Backhaus u.a. 2000, S. XVIII

³ Backhaus u.a. 2000, S. XVIII

⁴ <http://bernetblog.ch/2010/05/14/mediennutzung-neu-tv-vor-zeitung/>



Ein positives Beispiel für die sowohl gut verständliche als auch korrekte Darstellung von Ergebnissen ist das „Politbarometer“ der Forschungsgruppe Wahlen⁵. Eine schöne Zusammenstellung darüber, wie mit Grafiken „geschummelt“ werden kann und welche Angaben bei Grafiken wichtig sind, liefert das Statistische Landesamt mit einer anschaulichen Veröffentlichung⁶.

A3 Vollerhebung und Stichprobe

Vollerhebung:

Alle Personen/Objekte der relevanten Gruppe werden untersucht (z. B. alle SchülerInnen einer Klasse werden befragt)

Stichprobe:

Ein Teil der Personen/Objekte einer relevanten Gruppe („**Grundgesamtheit**“) wird untersucht (z. B. 10% aller Produkte eines Unternehmens werden geprüft)

Die meisten Datenanalysen beziehen sich auf Stichproben, also auf Daten von einem Teil der interessierenden Personen bzw. Objekte. Bei der Datenanalyse gibt es dann unterschiedliche Möglichkeiten.

Deskriptive Statistik:

Es werden Ergebnisse präsentiert, die sich nur auf die untersuchten Personen bzw. Objekte beziehen (z. B. „Die SchülerInnen der Klasse 3b finden ihre Klassenlehrerin toll“).

Schließende (induktive) Statistik:

Es werden Schätzungen präsentiert, bei denen aus den Ergebnissen der Stichprobe auf Eigenschaften von Personen bzw. Objekten in der „Grundgesamtheit“ geschlossen wird (z. B. „Aus der Stichprobe hat sich ergeben, dass ca. 3% aller Produkte der Baureihe XY fehlerhaft sind“).

Voraussetzung für die schließende Statistik ist, dass die Stichprobe „**repräsentativ**“ für die Grundgesamtheit ist, also eine ähnliche Verteilung bestimmter Merkmale aufweist (z. B. sollten bei einer Stichprobe von Produkten nicht nur Produkte ausgewählt werden, die von einer bestimmten Person produziert wurden).

A4 Merkmale und Merkmalsträger

Merkmalsträger:

Personen oder Objekte, die untersucht werden (z. B. SchülerInnen, Arbeitslose, Kühlschränke, Versuchslauf, Tag).

⁵ <http://www.forschungsgruppe.de/Aktuelles/Politbarometer/>

⁶ <http://www.statistik.baden-wuerttemberg.de/Veroeffentl/802011002.pdf>

**Merkmal** bzw. **Variable**⁷:

Eigenschaften, die ein Merkmalsträger haben kann (z. B. Gewicht, Dichte, Wochentag, Begeisterungsfähigkeit).

Merkmalsausprägung:

Mögliche Werte einer Eigenschaft bzw. eines Merkmals (z. B. 30 Gramm, Dienstag, hoch)

Wenn bei einer Untersuchung „etwas“ überprüft werden soll, muss also immer angegeben bzw. bekannt sein, um welche Merkmalsträger (z. B. SchülerInnen) und welche Merkmale (z. B. Gewicht) es gehen soll, und welche möglichen Merkmalsausprägungen erfasst werden sollen (z. B. „unter 20 kg, zwischen 21 und 40 kg, über 40 kg“).

A5 Skalen

Variablen (Merkmale) lassen sich unterschiedlich gut in Zahlen und damit auf einer **Skala**⁸ ausdrücken. Manche Variablen sind schon Zahlen (z. B. das Gewicht), manche sind gut in Zahlen übersetzbar und haben eine Reihenfolge (z. B. von 1=sehr schön bis 6=sehr hässlich) und manche kann man zwar mit einer Zahl **codieren**, jedoch ohne Reihenfolge (z. B. 1=rot, 2=blau, 3=braun).

Nominalskala (nominal skaliert):

Eigenschaften ohne Reihenfolge (z. B. Geschlecht, Farbe, Lieblingssendung). Eine Sonderform ist eine **dichotome** Variable, die genau zwei Ausprägungen hat (z. B. Ja/Nein).

Ordinalskala (ordinal skaliert):

Eigenschaften mit Reihenfolge, aber ohne mathematisch bestimmbare Abstände zwischen den einzelnen Werten (z. B. Zustimmung von 1=gar nicht bis 10=sehr).

Intervallskala (intervallskaliert):

Eigenschaften mit Reihenfolge und gleichen Abständen (z. B. Größe, Temperatur).

Ratio- oder Verhältnisskala (verhältnisskaliert):

Eigenschaften, bei denen der Nullpunkt bedeutet, dass es dort auch nichts gibt (z. B. kein Gewicht, keine Geschwindigkeit).

Die Intervallskala sowie die Ratio-/Verhältnisskala werden unter dem Begriff „**Metrische Skala**“ oder „Kardinalskala“ bzw. „**metrisch skaliert**“ zusammengefasst.

⁷ http://de.wikipedia.org/wiki/Statistische_Variable

⁸ Backhaus u.a. 2000, S. XIX



Davon, mit welcher Art von Skala eine Variable gemessen werden kann, hängt ab, welche Arten von Berechnungen mit der Variablen gemacht werden können (z. B. kann kein Mittelwert aus Lieblingsfarben berechnet werden).

A6 Datenmatrix

Die Daten einer Messung oder Befragung zu den Merkmalen, Merkmalsträgern und Merkmalsausprägungen werden in einer **Datenmatrix** gesammelt. Dabei stehen in der ersten Spalte (also links) die einzelnen Merkmalsträger und in der ersten Zeile (also oben) die einzelnen Merkmale.

	Klasse	Geschlecht	Größe	Interesse	Haarfarbe
Anna	3a	w	155	hoch	blond
Basti	3b	m	145	mittel	braun
Clara	3a	w	144	gering	schwarz
Daniel	3c	m	149	mittel	blond
Eva	3b	w	151	mittel	rot

Codierung:

Die Codierung „übersetzt“ alle Eigenschaften von Merkmalen in Zahlen. Bei Merkmalen mit Intervall- oder Ratioskala ist das nicht nötig. Merkmale mit Ordinal- oder Nominalskala jedoch brauchen eine festgelegte Zuordnung von Eigenschaften und Zahlenwerten.

	<i>Klasse</i>	<i>Geschlecht</i>	<i>Größe</i>	<i>Interesse</i>	<i>Haarfarbe</i>
<i>Anna</i>	1	1	155	3	2
<i>Basti</i>	2	0	145	2	3
<i>Clara</i>	1	1	144	1	4
<i>Daniel</i>	3	0	149	2	2
<i>Eva</i>	2	1	151	2	1

Fehlende Werte:

Wenn bei einer Befragung oder Messung ein Wert fehlt, muss in der Datenmatrix dennoch ein Wert angegeben werden. Hierfür wählt man dann ein Zeichen oder einen Wert, der bei den Eigenschaften der Variablen nicht vorkommen kann, meist eine 9, 99 oder 999.



Überblick / Einordnung

Typische Verfahren und Kennzahlen

	Deskriptive Statistik	Induktive Statistik
Univariate Statistik	Häufigkeiten, Mittelwerte, Streuungen: z. B. AM, Varianz	Schätzung: z. B. Konfidenzintervall des Arithmetischen Mittels
Bivariate Statistik	Kreuztabellen, Zusammenhänge: z. B. Pearson, η^2	Tests: z. B. T-Test, ANOVA
Multivariate Statistik	Homogenität, Zusammenhänge: z. B. Reliabilität, Regressionsanalyse	