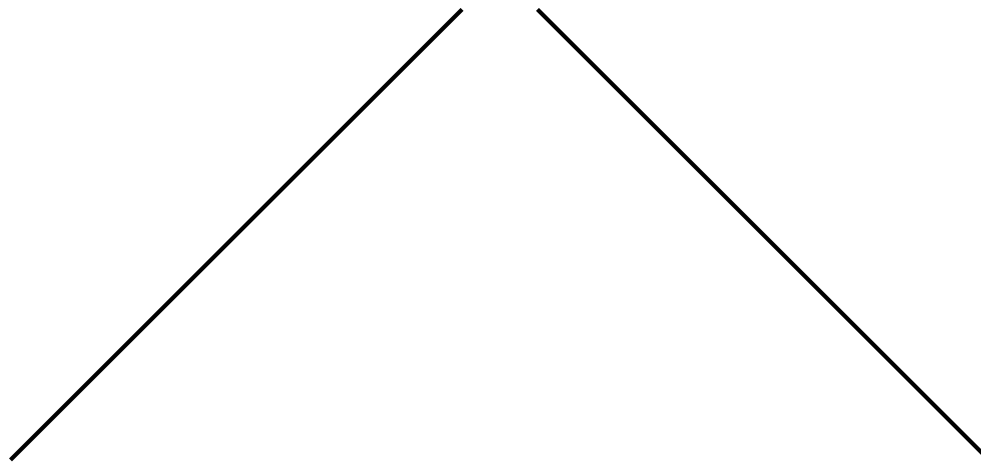


Statistik

Mathematische Hilfswissenschaft

mit der Aufgabe, Methoden für die Sammlung, Aufbereitung, Analyse und Interpretation von numerische Daten bereitzustellen, um die Struktur von Massenerscheinungen zu erkennen.



Deskriptive (beschreibende)

Statistik

- verfolgt das Ziel, die Untersuchungsergebnisse lediglich darzustellen (durch Tabellen, graph. Darstellungen, mit Charakteristischen Maßzahlen)

Schließende (analytische)

Statistik

(Inferenzstatistik)

- ihr Hauptanliegen ist es, von einer relativ kleinen Auswahl der Gesamtheit auf die Gesamtheit zu schließen (Repräsentationsschluss)
-> Induktionsschluss

Deskriptive Statistik

Statistische Beschreibung einer definierten Menge von Untersuchungseinheiten bezüglich einer Menge von Merkmalen (Eigenschaften).

Untersuchungseinheit: Die Untersuchungseinheit ist als Merkmalsträger das Bezugsobjekt der Sozialforschung (Einheit, auf die sich die Untersuchung bezieht).

z.B. Personen, Schulen, Texte

Merkmale: Die Eigenschaften der Untersuchungseinheiten

z.B. Lebensalter, Interessen, Einkommen

Merkmalsausprägungen: Werte, die die Merkmale annehmen

Beispiele: Merkmal: *Geschlecht*

Ausprägungen: *weiblich, männlich*

Merkmal: *Parteizugehörigkeit*

Ausprägungen: *SPD, CDU, FDP, Grüne, sonstige*

Merkmal: *Nettoeinkommen*

Ausprägungen: *1500 €, 2000 € etc.*

Variablen

Die Eigenschaften, die von Untersuchungseinheit zu Untersuchungseinheit variieren, d.h. verschiedenen Werte annehmen können, werden **VARIABLEN** genannt.

Variablen sind also:

- Eigenschaften, die man in Verbindung mit der Untersuchungseinheit erforscht
- Eigenschaften, die bei verschiedenen Untersuchungseinheiten in mehr als einer Ausprägung vorkommen

Beispiel: Schulklasse

Untersuchungseinheiten Variable Ausprägungen

Schüler 1 (x_1)		169
Schüler 2 (x_2)	Körperlänge	178
Schüler 3 (x_3)	in cm	174
...		...
...		...
Schüler n (x_n)		160

Menge

Beispiele:

Beschreibung aller Einwohner Essens über 18 Jahre hinsichtlich ihrer Zufriedenheit mit der gegenwärtigen Bundesregierung.

Beschreibung aller Besucher dieser Lehrveranstaltung hinsichtlich der Variablen „Alter“, „Anzahl der Semester“, „Geschlecht“, „Berufsziel“, usw.

Aufgabe:

Geben Sie für die folgende Fragestellung Untersuchungseinheiten; Variablen und Merkmalsausprägungen an:

Es soll die Todesursache von Rauchern ermittelt werden.

Untersuchungseinheiten:

Variable:

Merkmalsausprägungen:

Variablen

qualitativ

Unterscheidung
im Hinblick
auf ihre Art
z.B. Konfession

quantitativ

können verschiedene
Größen annehmen
und ihre Ausprägungen
können der Größe nach
geordnet werden

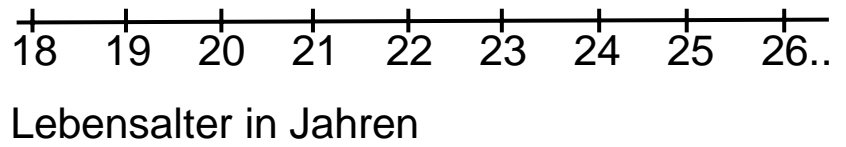
diskret

können nur endlich
viele oder abzählbar
unendlich viele Werte
annehmen
z.B.: Anzahl Einwohner
einer Stadt

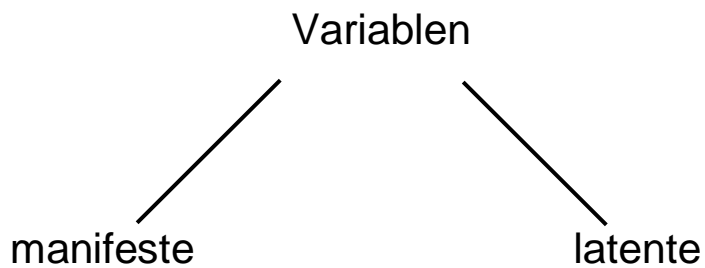
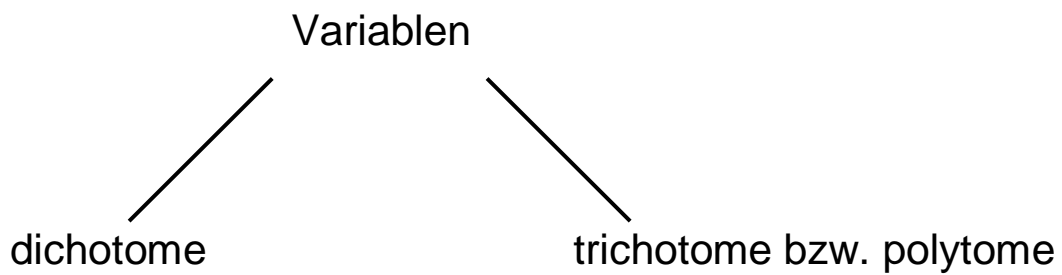
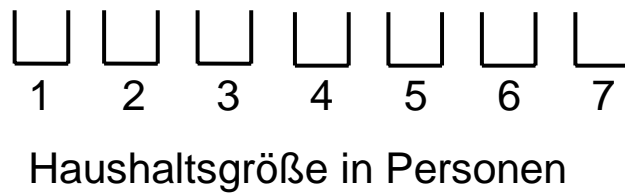
stetig

können jeden
beliebigen
Wert aus der
Menge der
reellen
Zahlen
annehmen
z.B.: Körper-
größe

stetige Variablen



diskrete Variablen



Der Stellenwert der Statistik in der sozialwissenschaftlichen Forschung

Zu den wichtigen Aktivitäten eines Sozialforschers gehören:

- 1) Die Beschreibung von Untersuchungseinheiten im Hinblick auf einzelne Variablen (z.B. Nettoeinkommen in der deutschen Bevölkerung)
- 2) Die Beschreibung der Beziehung zwischen Variablen (z.B. „Zufriedenheit mit der gegenwärtigen Berufsposition“ und „Anzahl Krankheitstage im Jahr“)
- 3) Die Generalisierung von Beobachtungsergebnissen

Tabelle 1: Urliste: Ungeordnete Auflistung der Variablenwerte

Variable: Lebensalter

44	28	47
31	21	44
32	15	43
30	43	37
56	44	31
57	47	30
44	82	29
45	61	28
44	60	18
29	60	19

Tabelle 2: Primäre Tafel: geordnete Auflistung der

Variablenwerte

Variable: Lebensalter

15	31	44
18	31	45
19	32	47
21	37	47
28	43	56
28	43	57
29	44	60
29	44	60
30	44	61
30	44	82

Tabelle 3: Häufigkeitstabelle: Die Verteilung der Variablen
Lebensalter

Messwert x_1	Strichmarke	Häufigkeit f_i
15		1
18		1
19		1
21		1
28		2
29		2
30		2
31		2
32		1
37		1
43		2
44		5
45		1
47		2
56		1
57		1
60		2
61		1
82		1
Gesamt		N = 30

Tabelle: Die Gewichtsverteilung einer Studentengruppe im Kurs Statistik

x_i	f_i	f_n	$\% f_i$	cum f_i	cum f_n	cum $f_i \%$
48	1	0,0137	1,37	1	0,0137	1
49	1	0,0137	1,37	2	0,0274	3
50	6	0,0822	8,22	8	0,1096	11
51	1	0,0137	1,37	9	0,1233	12
52	2	0,0274	2,74	11	0,1507	15
53	3	0,0411	4,11	14	0,1918	19
54	1	0,0137	1,37	15	0,2055	21
55	1	0,0137	1,37	16	0,2192	22
56	3	0,0411	4,11	19	0,2603	26
57	3	0,0411	4,11	22	0,3014	30
58	1	0,0137	1,37	23	0,3151	32
59	4	0,0548	5,48	27	0,3699	37
60	5	0,0685	6,85	32	0,4384	44
61	2	0,0274	2,74	34	0,4658	47
...
83	2	0,0274	2,74	72	0,9864	99
84	1	0,0137	1,37	73	1,0000	100
Σ	73	1,000	100			

N = 73

$$f_n = \frac{f_i}{n}$$

$$\% f_i = \frac{f_i}{n} \times 100$$

$$\text{cum } f_n = \frac{\text{cum } f_i}{n} = f_{c_i}$$

$$\text{cum } \% f_i = \frac{\text{cum } f_i}{n} \times 100$$

N bzw. n	Anzahl der Untersuchungseinheiten
X, Y, Z	Bezeichnung für Variablen
x, y, z	Werte, die die Variablen annehmen
$x_1, x_2, x_3 \dots x_n$	Werte der Variablen werden durch Indizes voneinander unterschieden
x_i	durch dieses Symbol werden die Werte der Variablen vertreten (wobei $i = 1, 2, 3, \dots, n$)
f_i	Häufigkeit des i-ten Werts (frequency)
Σ	besagt, dass zu addieren ist, was rechts neben ihm steht. Unterhalb und oberhalb des Summenzeichens wird die Summierungsgrenze angegeben. <u>Beispiel</u> : Die Summe der Werte $x_1+x_2+x_3+x_4$ kann wie folgt geschrieben werden:

$$\sum_{i=1}^4 x_i$$

allgemein gilt:

$$\sum_{i=1}^4 x_i = x_1 + x_2 + \dots + x_n$$

oder verkürzt:

$$\sum x_i$$

cum f_i	kumulierte absolute Häufigkeit	f_{c_i}
% f_i	prozentuale Häufigkeit	
f_n	relative Häufigkeit	
cum f_n	kumulierte relative Häufigkeit	f_{c_n}

Tabelle: Die Gewichtsverteilung einer Studentengruppe im Kurs Statistik

x_i	f_i	f_n	$\% f_i$	$\text{cum } f_i$	$\text{cum } f_n$	$\text{cum } f_i \%$
48	1	0,0137	1,37	1	0,0137	1
49	1	0,0137	1,37	2	0,0274	3
50	6	0,0822	8,22	8	0,1096	11
51	1	0,0137	1,37	9	0,1233	12
52	2	0,0274	2,74	11	0,1507	15
53	3	0,0411	4,11	14	0,1918	19
54	1	0,0137	1,37	15	0,2055	21
55	1	0,0137	1,37	16	0,2192	22
56	3	0,0411	4,11	19	0,2603	26
57	3	0,0411	4,11	22	0,3014	30
58	1	0,0137	1,37	23	0,3151	32
59	4	0,0548	5,48	27	0,3699	37
60	5	0,0685	6,85	32	0,4384	44
61	2	0,0274	2,74	34	0,4658	47
...
83	2	0,0274	2,74	72	0,9864	99
84	1	0,0137	1,37	73	1,0000	100
Σ	73	1,000	100			

N = 73

$$f_n = \frac{f_i}{n}$$

$$\% f_i = \frac{f_i}{n} \times 100$$

$$\text{cum } f_n = \frac{\text{cum } f_i}{n} = f_{c_i}$$

$$\text{cum } \% f_i = \frac{\text{cum } f_i}{n} \times 100$$

Tabelle: Die klassifizierte Gewichtsverteilung einer Studentengruppe im Kurs Statistik

Klassen-intervall	Math. Grenzen	Klassen-mittelwert	f_i	f_n	$\%f_i$	cum f_n	cum f_i %
46-50	45,5-50,49	48	8	0,11	11	0,11	11
51-55	50,0-55,49	53	8	0,11	11	0,22	22
56-60	55,5-60,49	58	16	0,22	22	0,44	44
61-65	60,5-65,49	63	14	0,19	19	0,63	63
66-70	65,5-70,49	68	9	0,12	12	0,75	75
71-75	70,5-75,49	73	12	0,16	16	0,92	92
76-80	75,5-80,49	78	2	0,03	3	0,95	95
81-85	80,5-85,49	83	4	0,05	5	1,00	100
			73	1,00	100		

Beispiele:

Univariate Häufigkeitsverteilung:

Geschlecht	Häufigkeit
Männer	72
Frauen	31
Σ (Summe)	103

Bivariate Häufigkeitsverteilung:

	Interesse an Statistik			
Geschlecht	viel	wenig	gar nicht	Σ
Männer	2	12	24	38
Frauen	22	5	3	30
Σ (Summe)	24	17	27	68