

Statistik

eine unvollständige Einführung

Warum Statistik?

- Weil Statistik zum Alltag dazugehört
- Weil Statistik interessant ist
- Und weil man mit Statistik gut lügen kann

	Erhebung 1	Erhebung 2
Frauen	2.8	2.8
Männer	2.2	2

Warum Statistik?

- Weil Statistik zum Alltag dazugehört
- Weil Statistik interessant ist
- Und weil man mit Statistik gut lügen kann

	Erhebung 1	Erhebung 2
Frauen	2.8	2.8
Männer	2.2	2
Gesamt	2.4	2.5

Warum Statistik?

- Weil Statistik zum Alltag dazugehört
- Weil Statistik interessant ist
- Und weil man mit Statistik gut lügen kann

	Erhebung 1		Erhebung 2	
		N		N
Frauen	2.8	10	2.8	20
Männer	2.2	20	2	10
Gesamt	2.4	30	2.5	30

Was ist Statistik?

- Deskriptive Statistik
 - Wie können wesentliche Gesichtspunkte eines Themas knapp und eindeutig beschrieben werden?
- Inferenzstatistik
 - Wenn eine Stichprobe vorliegt – welche Aussagen können über die Grundgesamtheit gemacht werden?

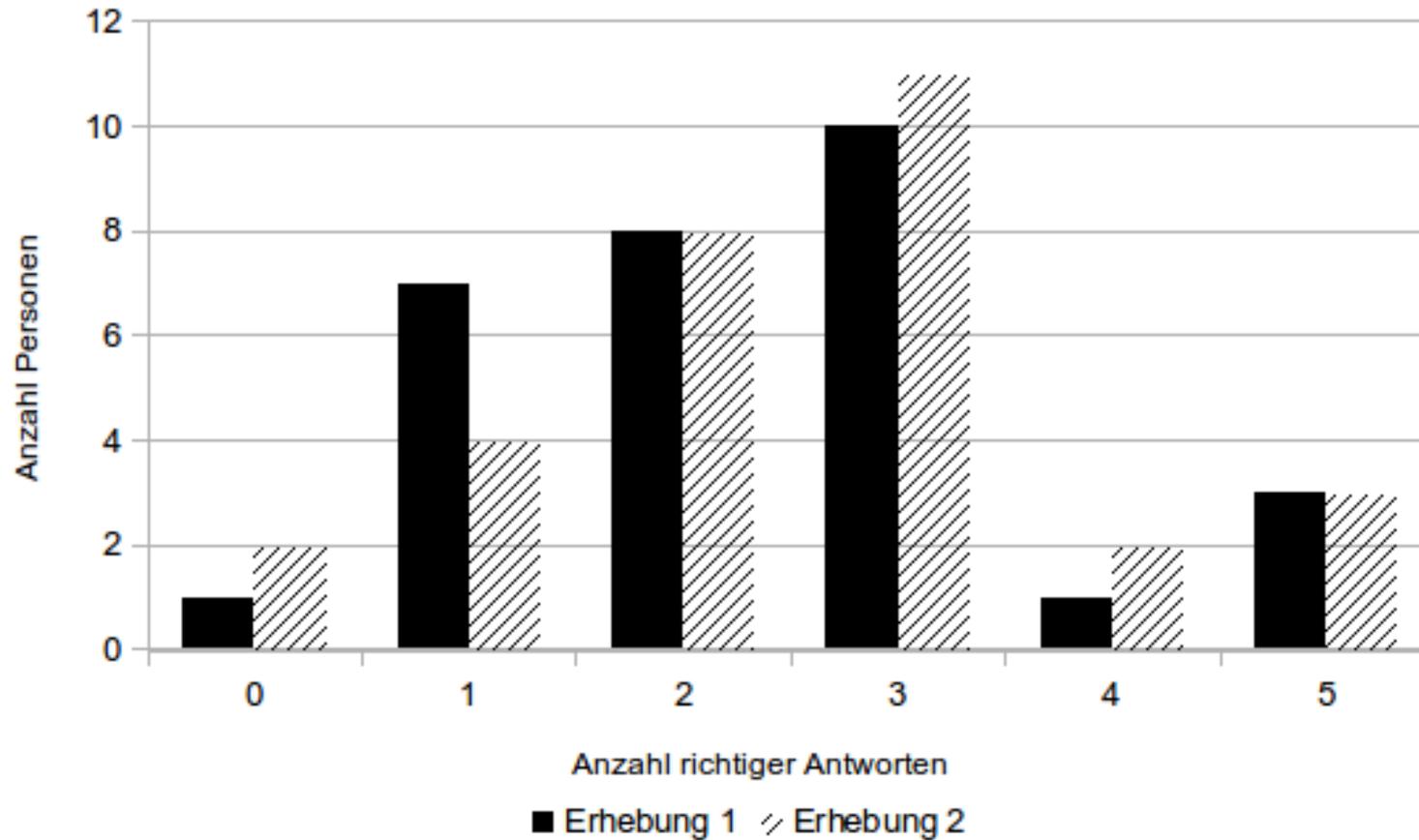
Notation

- N = Anzahl der Untersuchungseinheiten
- X = Variable
- I = einzelne, aber keine bestimmte Untersuchungseinheit $i \in (1, \dots, n)$
- X_i = die Ausprägung der Untersuchungseinheit i der Variable X

Deskriptive Statistik

Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frau	Erhebung	Aggr	
1	1	1	1	0	0	1	1	3
0	0	0	0	1	1	1	1	2
1	1	1	1	0	0	1	1	3
1	0	1	0	0	1	1	1	3
0	1	0	0	1	0	1	1	2
0	0	0	0	0	1	1	1	1
1	1	1	1	1	1	1	1	5
1	1	0	0	0	1	1	1	3
0	0	1	1	1	0	1	1	2
1	1	1	1	0	1	1	1	4
1	1	1	1	1	1	2	1	5
1	1	1	1	1	1	2	1	5
0	0	0	1	1	1	2	1	2
1	1	1	1	0	0	2	1	3
1	0	0	0	0	1	2	1	2
0	0	1	1	1	1	2	1	3
1	0	1	0	0	1	2	1	3
0	0	1	0	0	0	2	1	1
0	0	1	0	0	0	2	1	1
0	1	0	0	0	1	2	1	2
1	0	0	0	0	0	2	1	1
0	1	1	0	0	0	2	1	2
1	1	0	1	1	0	2	1	3
0	1	1	0	0	1	2	1	3
0	0	0	0	0	0	2	1	0
0	0	0	0	0	1	2	1	1
1	0	0	0	0	0	2	1	1

Deskriptive Statistik



Wichtige Maße

- Häufigkeiten
- Lagemaße
 - Modus
 - Median
 - Arithmetisches Mittel
- Streuungsmaße
 - Range
 - Varianz und Standardabweichung

Häufigkeiten

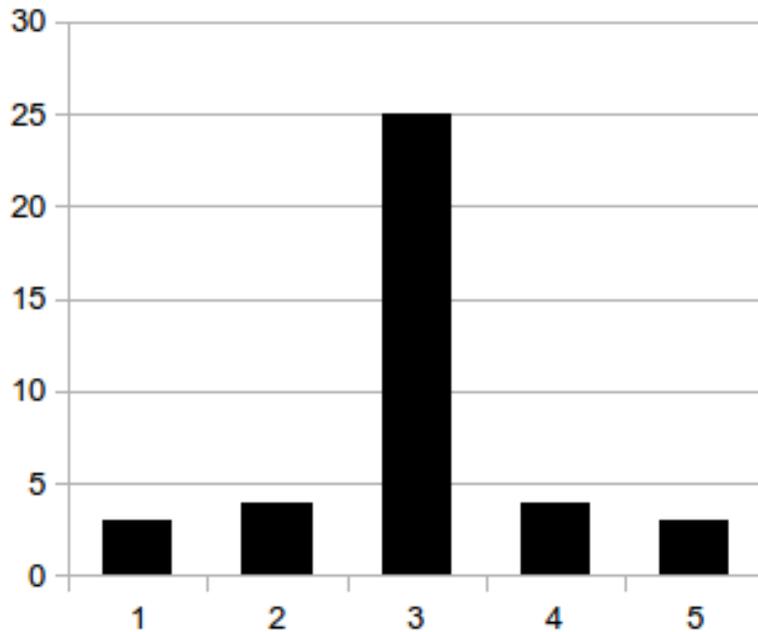
Ausprägung	Absolute Häufigkeit	Relative Häufigkeit (%)	Kumulierte relative Häufigkeit (%)
a1	h1	f1	F1
0	3	5.00	5.00
1	11	18.33	23.33
2	16	26.67	50.00
3	21	35.00	85.00
4	3	5.00	90.00
5	6	10.00	100.00
Gesamt	60	100.00	

Lagemaße

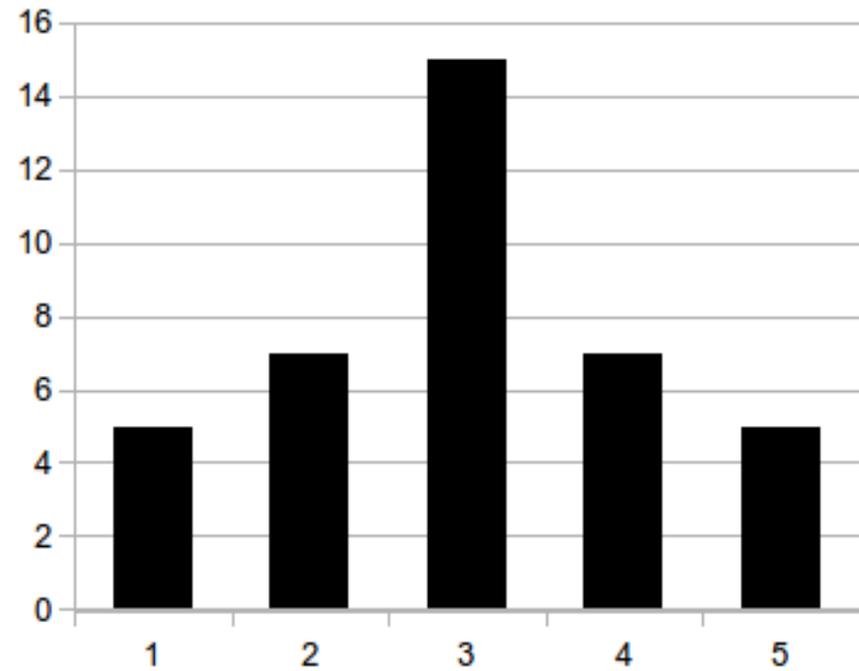
- Modus x_{MOD}
 - Die Ausprägung mit der grössten Häufigkeit
- Median x_{MED}
 - Der Wert, welcher die nach Grösse geordneten Daten genau in der Mitte teilt
- Arithmetisches Mittel
 - Die Summe der Einzelwerte dividiert durch deren Anzahl

$$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

Streuungsmaße



Modus: 3 Median: 3 arith. Mittel: 3



Modus: 3 Median: 3 Arith. Mittel: 3

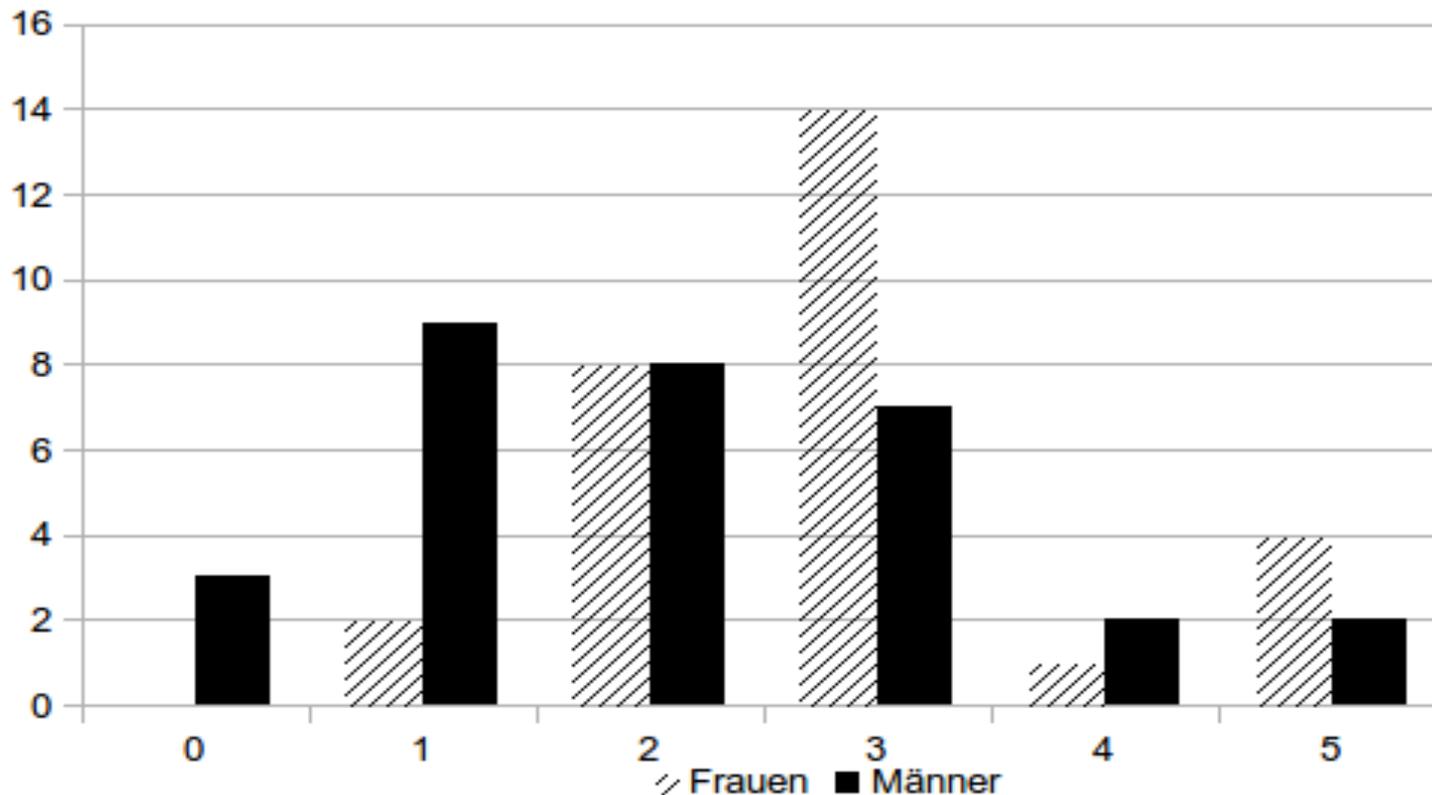
Streuungsmaße

- Range
 - Der Range R einer Verteilung ist der Abstand zwischen dem kleinstem und dem grösstem Wert
- Varianz
 - Durchschnittliche quadrierte Abweichung vom arithmetischem Mittel

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

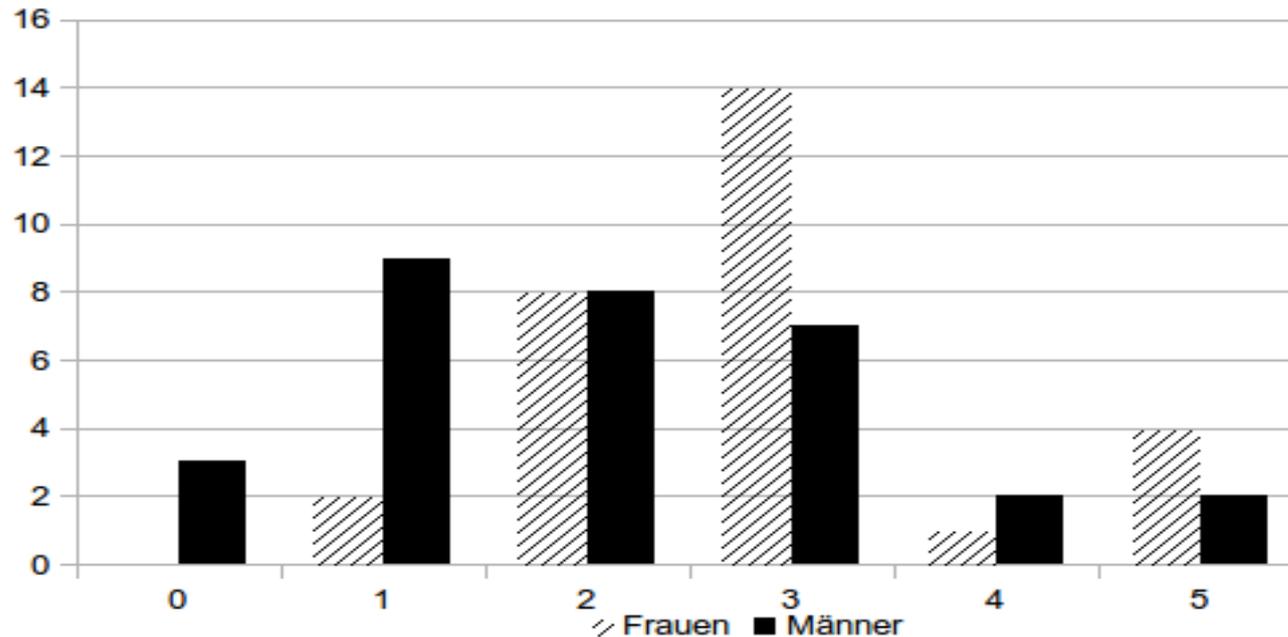
Bivariate Zusammenhänge

- Besteht zwischen zwei Variablen ein Zusammenhang?
- $X \rightarrow Y$ Hat X einen Einfluss auf Y ?



Bivariate Zusammenhänge

- Fragestellungen:
 - Besteht ein Zusammenhang zwischen X und Y?
 - In welche Richtung geht der Zusammenhang?
 - Wie stark ist der Zusammenhang?



Bivariate Zusammenhänge

- Zusammenhangsmaße
 - Kreuztabellen
 - Relatives Risiko
 - Odds-Ratio

Kreuztabellen

- Platziere X in den Spalten und Y in den Zeilen
- Berechne die Spaltenprozentante sowie die Randverteilungen
- Vergleiche die Prozentsätze in jeder Zeile

	Arbeiter		Angestellter		n
HS	360	90	270	45	630
RS/GM	40	10	330	55	370
n	400		600		1000

Kreuztabellen

Unabhängigkeit

		X	
		1	2
Y	1	40	40
	2	10	10
		50	50

sehr starker Zusammenhang

		X	
		1	2
Y	1	40	10
	2	10	40
		50	50

perfekter Zusammenhang

		X	
		1	2
Y	1	50	0
	2	0	50
		50	50

Relatives Risiko

- Das relative Risiko ist das Verhältnis zweier interessierender Anteilswerte
- Das Risiko von Arbeiterkindern, die Hauptschule zu besuchen ist mit $90\%/45\% = 2$ doppelt so gross wie das von Angestelltenkindern.

	Arbeiter	Angestellter	n
HS	90	45	630
RS/GM	10	55	370
n	400	600	1000

Odds-Ratio

- Die Odds-Ratio ist gegeben als das Verhältnis der bedingten Risiken (Odds), d.h. "Gewinnchancen", für $X = 1$ und $X = 2$
- Die Odds-Ratio also ist das Verhältnis der "Chancen" von Arbeiterkindern auf einen Hauptschulbesuch zuden "Chancen" von Angestelltenkindern auf einen Hauptschulbesuch

	Arbeiter	Angestellter	n
HS	360	270	630
RS/GM	40	330	370
n	400	600	1000
Odd	$360/40 = 9$	$270/330 = 0.818$	
Odds-Ratio	$9/0.818 = 11$		

Empfehlung:

<http://www.statisticshell.com/>



Quelle: Die Beispiele wurden teilweise der Vorlesung „Datenanalyse“ vom Prof. Thomas Gautschi an der Universität Mannheim (FSS2012) entnommen