

3-1

# STATISTIK FÜR ERZIEHUNGSWISSENSCHAFTLER:INNEN

## Vorlesung 3

Dr. Katja Serova



## 3-2 Rückblick

Urliste

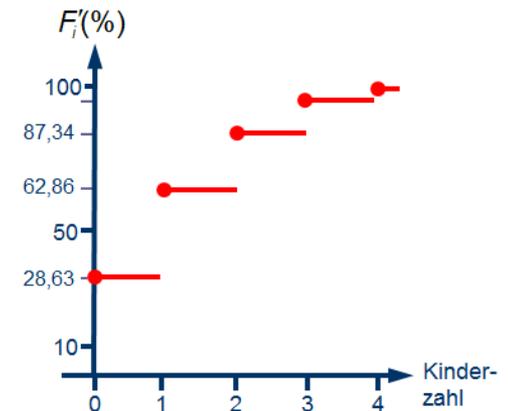
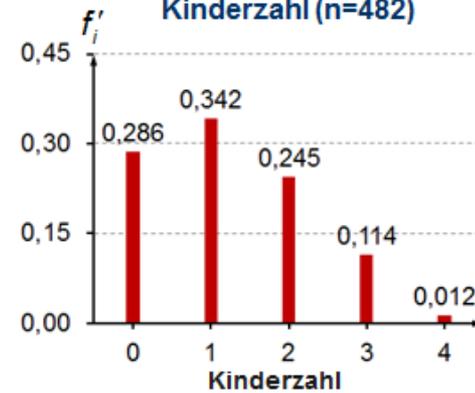
Nummer	Kinderzahl
141	0
142	1
143	1
144	0
145	2
146	0
147	2
148	3
149	1



Häufigkeitsverteilung

Kinderzahl $x_i$	Relative Häufigkeit (%) $f'_i$	Kumulierte Häufigkeit (%) $F'_i$
0	28,63 %	28,63 %
1	34,23 %	62,86 %
2	24,48 %	87,34 %
3	11,41 %	98,75 %
4	1,24 %	99,99 %
<b>Summe</b>	<b>99,99 % (*)</b>	

Häufigkeitsverteilung der Variablen  
Kinderzahl (n=482)



- Zusammenfassende Beschreibung und Darstellung einer **Datenmenge**
- **Überblick** über die gesamte Verteilung
- **Skalenniveau** und **Typ** der statistischen Variablen müssen beachtet werden

## 3-4 Deskriptive Statistik

Zusammenfassende **Beschreibung und Darstellung**  
einer **Datenmenge**

**ausführlich**

**Überblick** über eine Datenmenge

- tabellarisch
- graphisch

**konzentriert**

**Statistische Maßzahlen:** Darstellung  
ausgewählter Informationen über eine  
Datenmenge

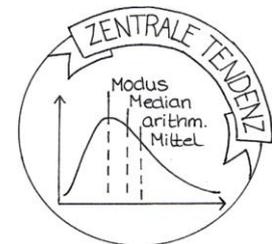
- Maße der zentralen Tendenz
- Streuungsmaße
- Zusammenhangsmaße
- Andere statistische Maße

3-5

# Bereich: Deskriptive Statistik

## Kapitel: Maße der zentralen Tendenz

- Was sind Maße der zentralen Tendenz?
- Bestimmung der Maße der zentralen Tendenz
- Unterschiedliche Maße der zentralen Tendenz im Vergleich



## 3-6 Einordnung

---

### Ziel

Zusammenfassende Beschreibung einer Datenmenge durch eine **statistische Maßzahl**

### Weg

Die **Mitte** finden

### Name

Maße der zentralen Tendenz, Mittelwerte, Lagemaße, Durchschnittswerte

### Arten

Modus, Median, arithmetisches Mittel, geometrisches Mittel, harmonisches Mittel etc.

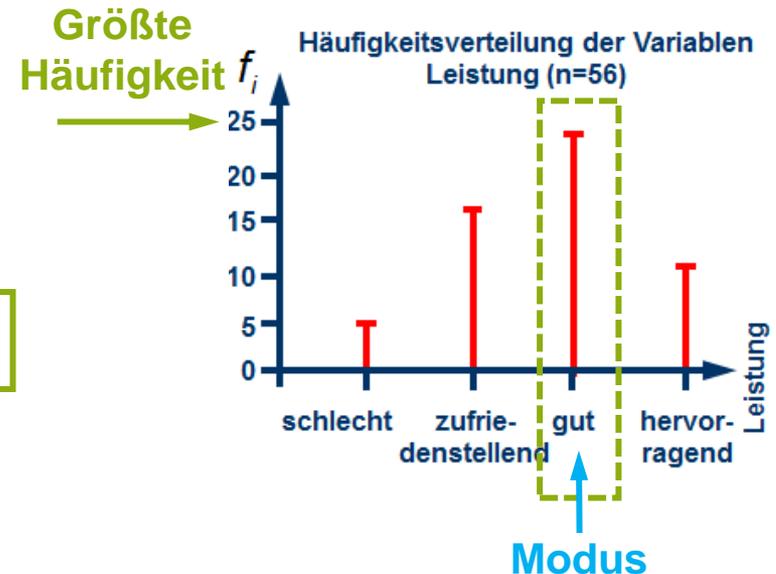
## 3-7 Modus (1)

- Die **häufigste Merkmalsausprägung** (der häufigste Wert)
- Lagetypisches Maß (nicht rechnerisch)
- Skalenniveau: nominal, ordinal und metrisch

Leistung $X_i$	$f_i$	$f'_i(\%)$
schlecht	5	8,93 %
zufriedenstellend	16	28,57 %
gut	24	42,86 %
hervorragend	11	19,64 %
Summe	56	100 %

Modus = gut

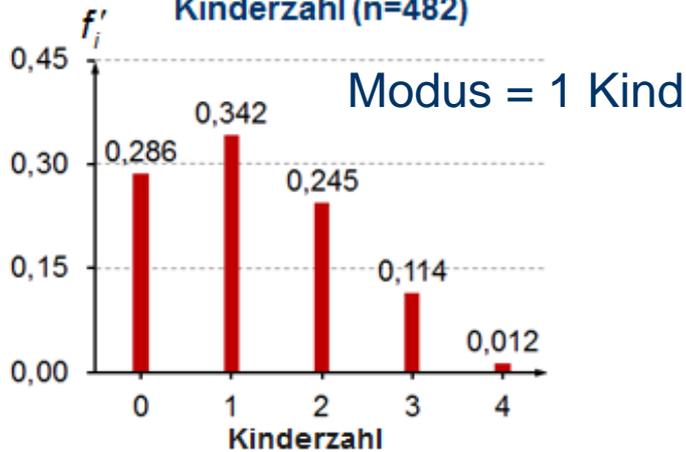
Größte Häufigkeit in der Verteilung



~~Modus = 24~~  
~~Modus = 42,86 %~~ **Unsinn!**

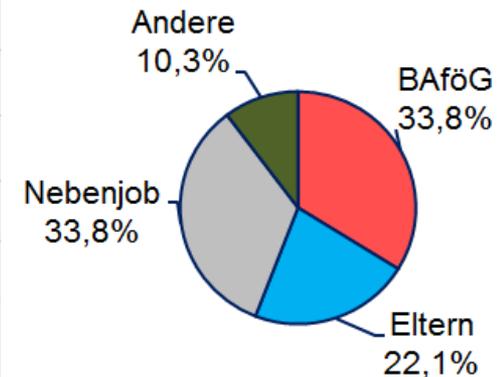
## 3-8 Modus (2)

Häufigkeitsverteilung der Variablen  
Kinderzahl (n=482)

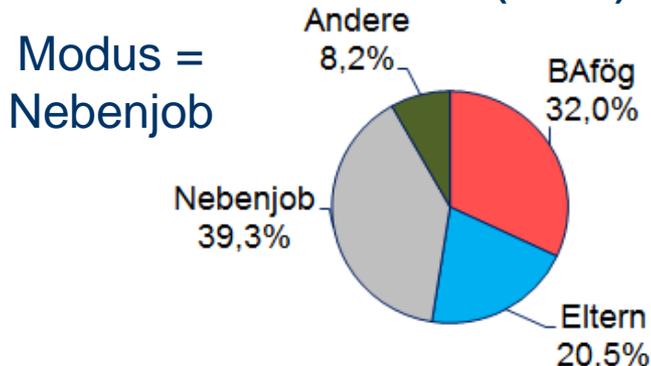


Hauptfinanzierungsquelle,  
EW-Studierende (n=68)

Hauptfinanzierungsquelle	EW-Studierende
BAföG	21
Eltern	15
Nebenjob	21
Andere	7
Summe	68



Hauptfinanzierungsquelle,  
SoWi-Studierende (n=122)



Modus = BAföG

Modus = Nebenjob

- Ein Modus: Unimodale Verteilung
- Zwei Modi: Bimodale Verteilung
- Drei und mehr Modi: Multimodale Verteilung

# ***Diese Folien (10-17) stehen Ihnen in dieser Version nicht zur Verfügung.***

An dieser Stelle wird in der Vorlesung die stetige Häufigkeitsverteilung (klassifizierte Daten) mit gleicher und ungleicher Klassenbreite sowie die Vor- und Nachteile des Modus vorgestellt. Anschließend folgt eine Einführung in den Median und die Urliste.

## 3-17 Median (5)

Eine **diskrete** Häufigkeitsverteilung

Leistung $x_i$	$f_i$	$F_i$
schlecht	5	5
zufriedenstellend	16	21
gut	24	45
hervorragend	11	56
Summe	56	

*Nummerierung*

- **Weg 1:** Formel und absolute kumulierte Häufigkeiten

$$\text{Median} = \frac{\frac{x_n}{2} + \frac{x_{n+2}}{2}}{2}$$

$$\text{Median} = \frac{x_{28} + x_{29}}{2}$$

**Median = gut**

## 3-18 Median (6)

Eine **diskrete** Häufigkeitsverteilung

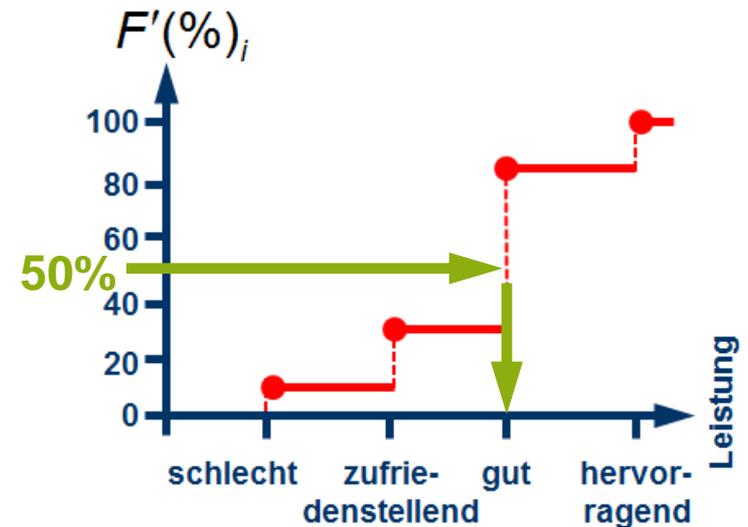
Leistung $x_i$	$f_i$	$F_i$	$f'_i(\%)$	$F'_i(\%)$
schlecht	5	5	8,93 %	8,93 %
zufriedenstellend	16	21	28,57 %	37,50 %
gut	24	45	42,86 %	80,36 %
hervorragend	11	56	19,64 %	100 %
Summe	56		100 %	

- **Weg 2:** Relative kumulierte Häufigkeiten

$Median = x_i \rightarrow$  bei  $F'_i = 50 \%$

**Median = gut**

- **Weg 3:** Graphisch aus der Treppenfunktion



**Median = gut**

## 3-19 Median (7)

Eine **stetige** Häufigkeitsverteilung (klassifizierte Daten)

Gewicht männlicher Studenten (kg)	$f_i$	$F_i$	$f'_i(\%)$	$F'_i(\%)$
50 bis unter 60	6	6	7,50 %	7,50 %
60 bis unter 70	17	23	21,25 %	28,75 %
70 bis unter 80	25	48	31,25 %	60,00 %
80 bis unter 90	20	68	25,00 %	85,00 %
90 bis unter 100	12	80	15,00 %	100,00 %
Summe	80		100,00 %	

- **Weg 1:** Formel und absolute kumulierte Häufigkeiten

$$\text{Median} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_{40} + x_{41}}{2} = 75$$

- **Weg 2:** Relative kumulierte Häufigkeiten

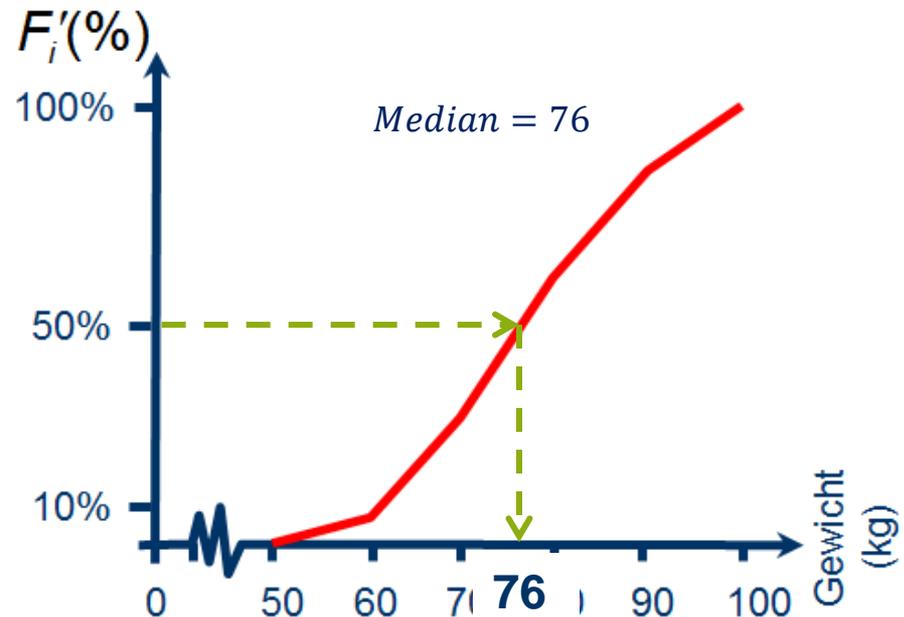
$$\text{Median} = x_i \rightarrow \text{bei } F'_i = 50 \%$$

$$\text{Median} = 75$$

Weg 3: Genauer als aus der Tabelle

Weg 1 und 2:  
Ohne  
Feinberechnung  
ungenau

- **Weg 3:** Graphisch aus der Ogive



***Diese Folien (21-22) stehen Ihnen in dieser Version nicht zur Verfügung.***

An dieser Stelle wird in der Vorlesung ein exemplarisches Beispiel sowie die Vor- und Nachteile des Median vorgestellt.

# 3-22 Arithmetischer Mittelwert (1)

- Für eine **Urliste**

**Variable:** Anzahl der Statistikbücher

**Messungen:** 1,3,0,1,2

**AM = ?**

Summe aller Bücher

$$\frac{1 + 3 + 0 + 1 + 2}{5} = \frac{7}{5} = 1,4$$

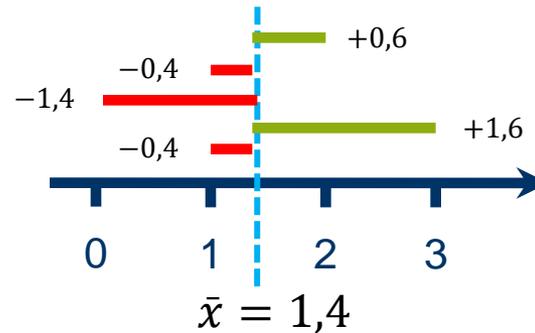
Anzahl der Messungen

- Formel: Einfacher arithmetischer Mittelwert**

*x quer*

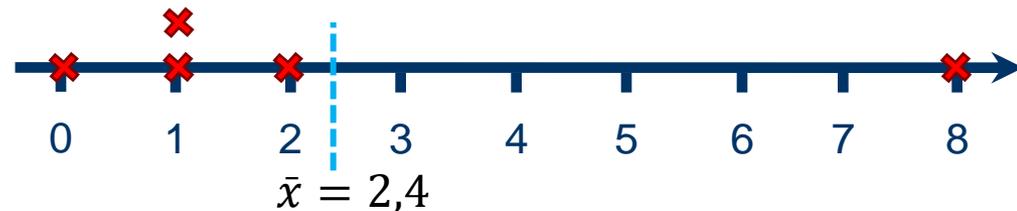
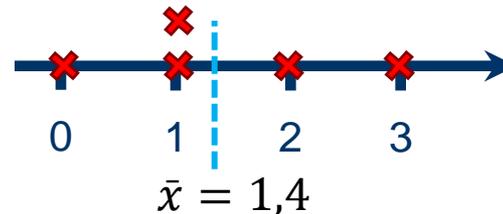
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- Abweichungen  $x_i - \bar{x}$



$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

- Extremwerte



## 3-23 Arithmetischer Mittelwert (2)

- Für eine **diskrete Häufigkeitsverteilung**

Kinderzahl $x_i$	$f_i$	$x_i f_i$
0	138	0*138 = 0
1	165	1*165 = 165
2	118	2*118 = 236
3	55	3*55 = 165
4	6	4*6 = 24
<b>Summe</b>	<b>482</b>	<b>590</b>

Anzahl der Familien

Gesamtanzahl der Kinder

$$\sum_{i=1}^k f_i = n = 482$$

$$\sum_{i=1}^k x_i f_i = 590$$

$$\bar{x} = \frac{\text{Gesamtanzahl der Kinder}}{\text{Anzahl der Familien}} = \frac{?}{482} = \frac{590}{482} \approx 1,22$$

- Formel: Gewogener (gewichteter) arithmetischer Mittelwert**

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i}$$

Gesamtanzahl der Kinder

Gesamtanzahl der Familien

## 3-24 Arithmetischer Mittelwert (3)

- Für eine stetige Häufigkeitsverteilung

Gewicht (kg)	Klassen- mitte $X_i$	$f_i$	$X_i f_i$
50 bis unter 60	55	6	55*6=330
60 bis unter 70	65	17	65*17=1105
70 bis unter 80	75	25	75*25=1875
80 bis unter 90	85	20	85*20=1700
90 bis unter 100	95	12	95*12=1140
Summe		80	6150

Gewogener (gewichteter) arithmetischer Mittelwert

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i} = \frac{6150}{80} = 76,875$$

Gewicht (kg)	Klassen- mitte $X_i$	$f_i$	$X_i f_i$
50 bis unter 70	60	23	60*23=1380
70 bis unter 100	85	57	85*57=4845
Summe		80	6225

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i} = \frac{6225}{80} = 77,8125 \approx 77,81$$

Nur die Bestimmung  
aus der Urliste ist  
genau!

***Diese Folie steht Ihnen in dieser Version  
nicht zur Verfügung.***

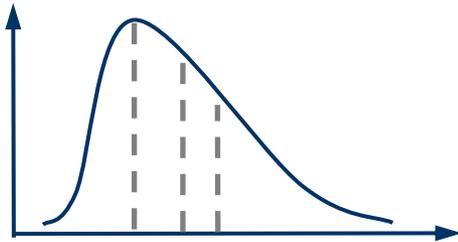
An dieser Stelle werden in der Vorlesung die Vor- und Nachteile des Arithmetischen Mittelwerts vorgestellt.

# Maße der zentralen Tendenz

## 3-26 im Vergleich (2)

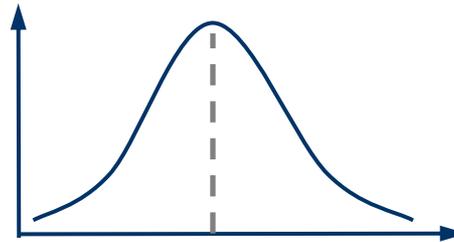
- Unterschiedliche **unimodale** Verteilungsformen

*linkssteil  
rechtsschief*



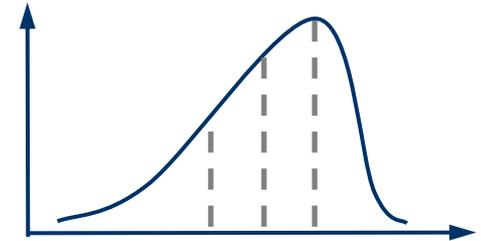
Modus < Median < AM

*symmetrisch*



Modus = Median = AM

*linksschief  
rechtssteil*



AM < Median < Modus

- Unterschiedliche **Skalenniveaus**

Skalenniveau	Modus	Median	AM
METRISCH	ja	ja	ja
ORDINAL	ja	ja	nein
NOMINAL	ja	nein	nein

***Diese Folie steht Ihnen in dieser Version  
nicht zur Verfügung.***

An dieser Stelle werden in der Vorlesung die Vor- und Nachteile des Arithmetischen Mittelwerts vorgestellt.

3-29

Nächste Übung: **Aufgabenblatt 03**

Nächste Vorlesung

**Bereich: Deskriptive Statistik**

**Kapitel: Streuungsmaße**



**Bis zum nächsten Mal  
oder bis Freitag!**

## 3-30 Texte zu diesem Kapitel



- **Mittelwerte** - aus Ehrenberg, Andrew S. C. (1986): Statistik oder der Umgang mit Daten. Eine praktische Einführungen mit Übungen. Weinheim: VCH, S. 3-13.
- **Summary Statistics & Median** - aus Gonick, Larry; Smith, Woollcott (1993): The Cartoon Guide To Statistics. New York: Harper, S.14-18.

### Texte zur Notation

- **Zum Gebrauch statistischer Formeln** - aus Buttler, Günther (2002): Einführung in die Statistik. Hamburg S.46-49.
- **Question: Notations** - Diamond, Ian; Jefferies, Julie (2001): Beginning Statistics. An Introduction for Social Scientists. London: Sage publications. S.238-247 (Auszüge).

### Zusätzliche Texte

- **Analyse eindimensionaler Häufigkeitsverteilungen** - aus Sauerbier, Thomas (2003): Statistik für Wirtschaftswissenschaftler. 2., überarbeitete Auflage. München: Oldenbourg, S.21-29 (Auszüge).

Quellen für die Bilder in den Folien:

- [www.freepik.com](http://www.freepik.com)
- [www.icon-icons.com](http://www.icon-icons.com)
- WAZ und ADAC