WIRTSCHAFTSSTATISTIK
MODUL 4: LAGEPARAMETER

WS 2023/24 DR. E. MERINS

LAGEPARAMETER

→ **noch Lokationsmaße genannt**

→ **Lageparameter beschreiben die “Lage” der Elemente der Grundgesamtheit bzw. der Stichprobe in Bezug auf die Messskala**

ALLGEMEINE LAGEPARAMETER

Allgemeine Mittelwerte:

- **Modus** $\bar{x}_D$
- **Median** $\bar{x}_Z$
- **arithmetisches Mittel** $\bar{x}$
- **Quantil** $\tilde{x}_p$

Spezielle Mittelwerte werden wir in diesem Kurs nicht berechnen. Sie müssen aber wissen, dass es solche existieren

Spezielle Mittelwerte:

- geometrisches Mittel $\bar{x}_G$
  (oder die mittlere Proportionale)
  wird in der Regel im Zusammenhang mit Wachstumsfaktoren wie Preissteigerungen oder Zinsen verwendet.
- harmonisches Mittel $\bar{x}_H$
  (ein Mittelwert einer Menge von Zahlen)
  wird verwendet, um den Mittelwert von Verhältniszahlen (Quotient zweier Größen) zu berechnen

MODUS

- **Modus (oder Modalwert)** $\overline{x}_D$

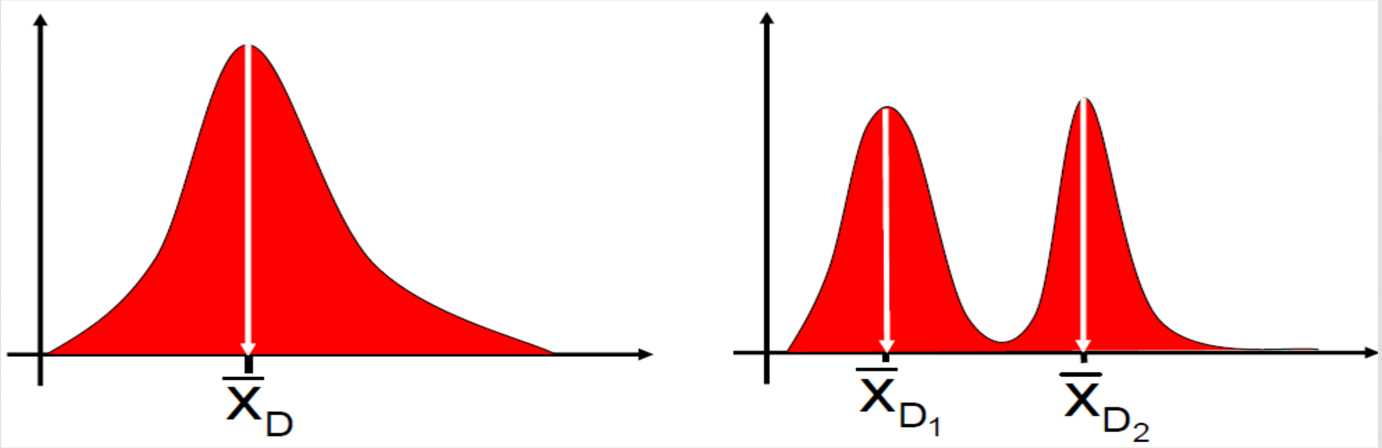
Der **Modus** oder **Modalwert** ist die am häufigsten auftretende Merkmalsausprägung (**maximale Häufigkeit**). Er wird hauptsächlich für nominale Merkmale verwendet, ist aber auch für alle anderen (diskreten) Merkmalstypen sinnvoll.

Bei klassierten Daten ist der Modalwert die **Mitte der Klasse mit den größten Häufigkeiten**. Diese Klasse nennt man die **Modalklasse**.

<u>Bemerkung:</u>
Gibt es mehrere Merkmalsausprägungen mit der gleichen maximalen Häufigkeit, so existieren <u>mehrere</u> Modalwerte → **Multimodale Verteilungen** (bimodale Verteilung: zwei Modalwerte; trimodale Verteilung: drei Modalwerte; usw.)

MODUS

- **Modus (oder Modalwert) $\overline{x}_D$:** Merkmalsausprägung, die am häufigsten vorkommt



unimodale Verteilung

Dichtekurve hat nur ein lokales Maximum

multimodale Verteilung

Dichtekurve hat mehrere lokale Maxima (bimodale Verteilung, Trimodale Verteilung usw.)

MEDIAN

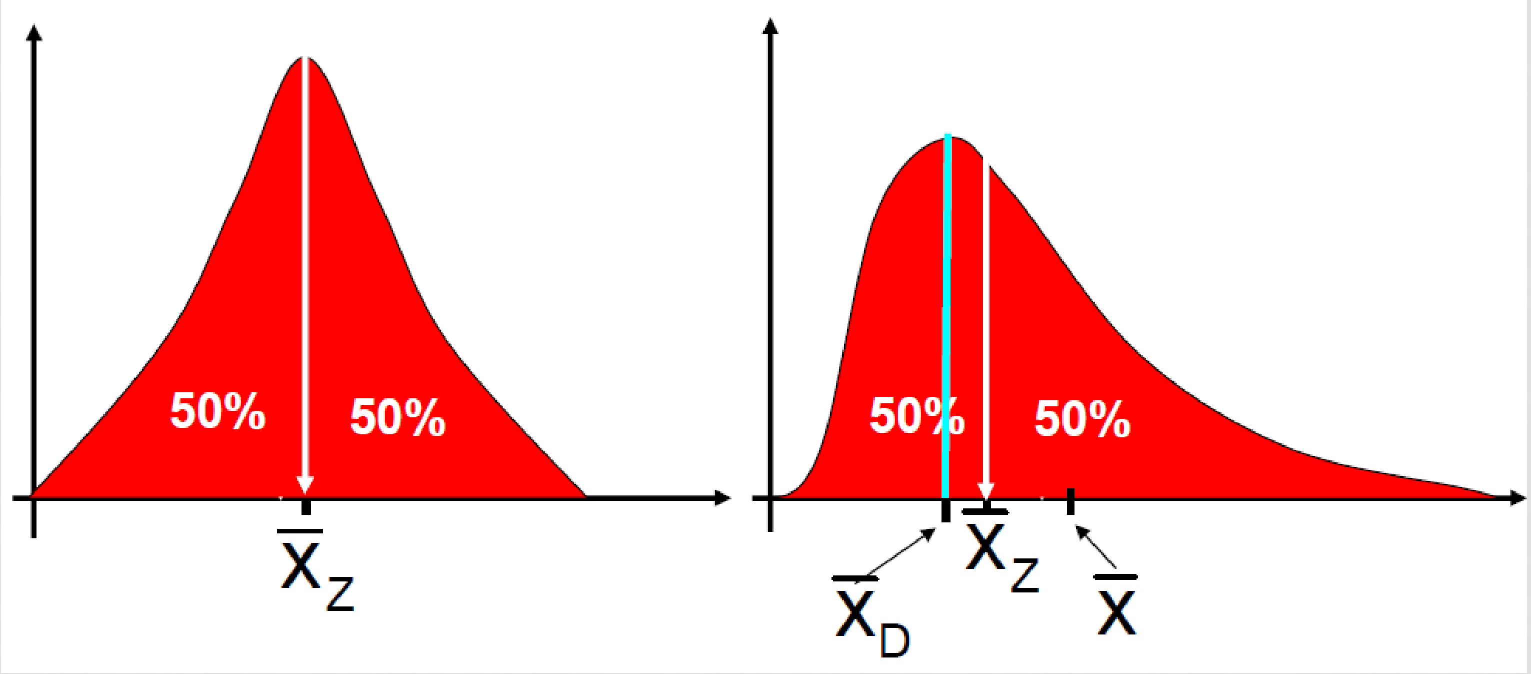
- **Median (oder Zentralwert)** $\bar{x}_z$

Mindestens 50% der Werte liegen links und mindestens 50% rechts des Medians (den Median selbst ggf. mit eingerechnet).

Median ist ein sehr robustes Lokationsmaß. Robuste statistische Kenngrößen sind wenig anfällig gegen Datenausreißer. Man muss die Hälfte der Daten gegen $+\infty$ oder $-\infty$ verschieben, um den Median selbst gegen $\pm\infty$ wandern zu lassen.

MEDIAN

- **Median (oder Zentralwert) $\bar{x}_Z$:** Wert in der Mitte der geordneten Reihe. 50% der Beobachtungswerte liegen unter dem Median, 50% darüber.



MEDIAN

- **Median (oder Zentralwert)** $\overline{x}_Z$

Für ordinale und metrische Merkmale ist der empirische Median (oder Zentralwert) definiert als:

$$\overline{x}_Z := x_{0,5} := \begin{cases} x_{\frac{n+1}{2}}, & falls\ n\ ungerade \\ \frac{1}{2} * \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right), & falls\ n\ gerade \end{cases}$$

MEDIAN

- **Median (oder Zentralwert)** $\bar{x}_Z$

Beispiel 1:

Die Anzahl n der Merkmalsausprägungen ist ungerade, z.B. das Alter von 7 Lehrern

(n = 7)

28	31	40	45	52	53	62
x_1	x_2	x_3	x_4	x_5	x_6	x_7
3 Werte			$\bar{x}_Z$	3 Werte		

$$\bar{x}_Z = x_{\frac{n+1}{2}} = x_{\frac{7+1}{2}} = x_4 = 45$$

In der Tabelle stehen links und rechts neben dem Median gleich viele Werte.

MEDIAN

- **Median (oder Zentralwert)** $\overline{x}_Z$

Beispiel 2:

Die Anzahl n der Merkmalsausprägungen ist gerade, z.B. das Alter von 8 Lehrern (n = 8)

28	31	40	45	52	53	58	62
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
3 Werte			$\overline{x}_Z$		3 Werte		

$$\overline{x}_Z = \frac{1}{2} * \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) = \frac{1}{2} * \left(x_{\frac{8}{2}} + x_{\frac{8}{2}+1}\right) = \frac{1}{2} * (x_4 + x_5) = \frac{1}{2} * (45 + 52) = 48{,}5$$

Bei einer geraden Anzahl von Werten berechnet man den Median aus den beiden mittleren Werten.

MEDIAN

- **Median (oder Zentralwert)** $\bar{x}_Z$

Bemerkungen:

Falls das betrachtete Merkmal nur ordinal skaliert ist (z.B. Zeugnisnoten), so ist bei geradem n zu beachten, dass der Median nur dann existiert, wenn beide infrage kommenden Merkmalsausprägungen gleich sind.

Beispiel:

bei den Zeugnisnoten 1 2 3 4 5 6 existiert kein Median, denn 3,5 als Zeugnisnote ist nicht üblich.

Aber: 1 2 3 3 4 5 hat den Median 3.

MEDIAN BEI KLASSIERTEN DATEN

- **Median (oder Zentralwert)** $\overline{\boldsymbol{x}}_{\boldsymbol{z}}$

Für metrische Daten in Klassen, kann die exakte Merkmalsausprägung des Medians nicht bestimmt werden → **Näherungswerte für Median**

→ **Näherungswert**, weil die Verteilung in den Klassen nicht bekannt ist. Es wird angenommen, dass die Beobachtungswerte in den Klassen gleich verteilt sind.

MEDIAN BEI KLASSIERTEN DATEN

Berechnung des Näherungswertes für Median:

$$\overline{x}_Z \coloneqq x_{k-1} + (x_k - x_{k-1}) * \frac{0{,}5 - F_{k-1}}{f_k}$$

k = Einfallsklasse → Klasse mit F(x)=50%

relative Summenhäufigkeit der Klasse Unterhalb der Einfallsklasse (Anteilswert)

$$\boxed{\overline{x}_Z = x^*_{k-1} + (x^*_k - x^*_{k-1}) \cdot \frac{(0{,}5 - F^*_{k-1})}{f_k}}$$

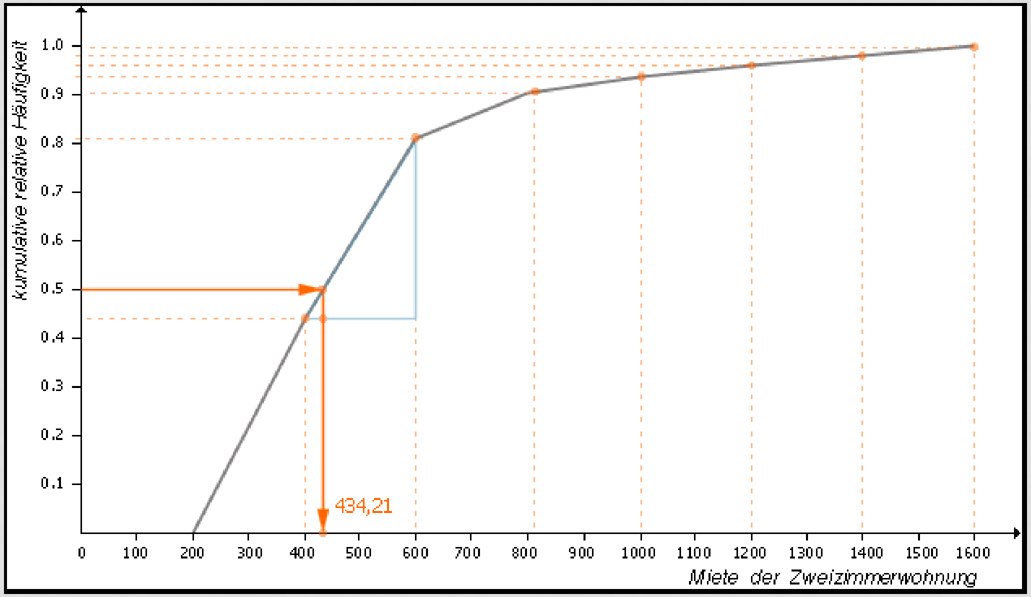
Untergrenze der Einfallsklasse

Obergrenze der Einfallsklasse

relative Häufigkeit der Einfallsklasse Anteilswert

MEDIAN BEI KLASSIERTEN DATEN

Beispiel 1:



Das Ergebnis:

<u>Näherungswert</u> für Median aus klassierten Daten $\bar{x}_Z$ = 434 €.

Der <u>tatsächliche</u> Median der Daten ist 423 €.

Achtung!! Es existiert Fehler der Näherung.

MEDIAN BEI KLASSIERTEN DATEN

Beispiel 2:

Klassen-Nr. i	Größen-klassen (cm)	h_i	f_i (%)	H_i	F_i (%)
1	100 b.u. 150	40	40%	40	40%
2	150 b.u. 170	40	40%	80	80%
3	170 b.u. 200	20	20%	100	100%
Summe		100	100%	-	-

Einfallsklasse: k = 2

$$\bar{x}_Z = x^*_{k-1} + (x^*_k - x^*_{k-1}) \cdot \frac{(0{,}5 - F_{k-1})}{f_k}$$

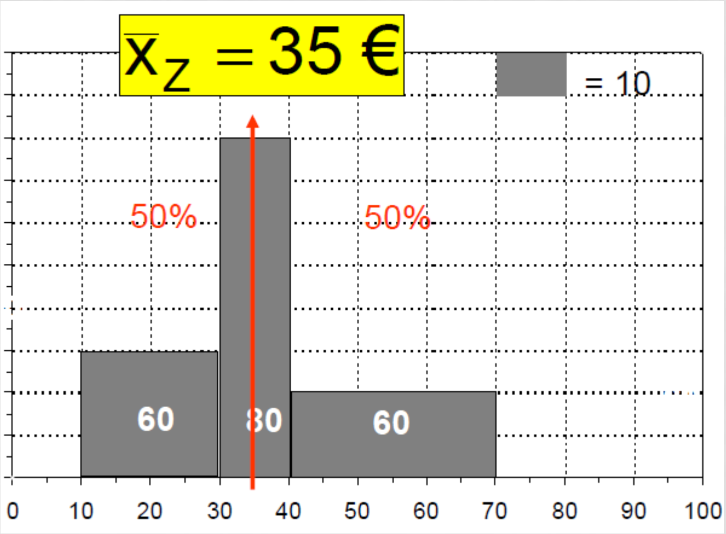
$$\bar{x}_Z = 150 + (170 - 150) \cdot \frac{(0{,}5 - 0{,}4)}{0{,}4} = 150 + 20 \cdot \left(\frac{0{,}1}{0{,}4}\right) = 155 \text{ (cm)}$$

MEDIAN BEI KLASSIERTEN DATEN

Beispiel 3:

Umsatz-klassen (€)	Anzahl h_i	Anteil in % f_i	H_i	F_i (%)
10 b.u. 30	60	30%	60	30%
30 b.u. 40	80	40%	140	70%
40 b.u. 70	60	30%	200	100%
Summe	200	100%	-	-

Klassen-Nr.
1
2
3



$$\bar{x}_Z = x^*_{k-1} + (x^*_k - x^*_{k-1}) \cdot \frac{(0{,}5 - F_{k-1})}{f_k}$$

Einfallsklasse k = 2 (30 b.u. 40)

$$\bar{x}_Z = \mathbf{30 + (40 - 30)\frac{0{,}5 - 0{,}3}{0{,}4} = 35} \text{ (€)}$$

ARITHMETISCHES MITTEL

- **Arithmetisches Mittel** $\bar{x}$

Das **arithmetische Mittel** (oder **Mittelwert**, oder **Durchschnitt** genannt) ist sinnvoll für beliebige <u>metrische</u> Merkmale.

$$\bar{x} = \bar{x}_n = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Statistiker-Witz:

Steht jemand mit einem Fuß auf der Herdplatte und mit dem anderen im Eiskasten, dann sagt der Statistiker: im Durchschnitt ist ihm angenehm warm.

ARITHMETISCHES MITTEL

- **Arithmetisches Mittel** $\overline{\boldsymbol{x}}$

Eigenschaften:

- Die Summe der Abweichungen der Einzelwerte vom arithmetischen Mittel ist stets gleich null $\sum(\boldsymbol{x_i} - \overline{\boldsymbol{x}}) = \mathbf{0}$
- bekanntester Mittelwert
- nur für quantitative Merkmale sinnvoll
- empfindlich gegen Ausreißer (Vorsicht bei schiefen Verteilungen!)

ARITHMETISCHES MITTEL AUS HÄUFIGKEITSTABELLEN

- **Arithmetisches Mittel** $\boldsymbol{\overline{x}}$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{j} x_i * h(x_i) = \sum_{i=1}^{j} x_i * f(x_i)$$

$x_1, \ldots, x_j$	***Merkmalsausprägungen***
$h(x_1), \ldots, h(x_j)$	***absolute Häufigkeiten***
$f(x_1), \ldots, f(x_j)$	***relative Häufigkeiten***

ARITHMETISCHES MITTEL AUS HÄUFIGKEITSTABELLEN

- **Arithmetisches Mittel** $\overline{x}$

Fall 1: Absolute Häufigkeit $h(x_i)$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{j} x_i * h(x_i) = \frac{1}{n} * (x_1 h(x_1) + x_2 h(x_2) + \cdots + x_j h(x_j))$$

$$n = \sum_{i=1}^{j} h(x_i) = h(x_1) + h(x_2) + \cdots + h(x_j)$$

$h(x_i)$ *absolute Häufigkeit der Merkmalsausprägung* x_i

n *Summe der absoluten Häufigkeiten*

j *Anzahl der Merkmalsausprägungen* x_i

ARITHMETISCHES MITTEL AUS HÄUFIGKEITSTABELLEN

- **Arithmetisches Mittel** $\overline{x}$

Beispiel:

Berechnung des arithmetischen Mittels über die absoluten Häufigkeiten:

Note x_i	1	2	3	4	5	6
Anzahl Schüler $h(x_i)$	5	8	14	16	5	2

Schüler insgesamt:

$$n = \sum_{1}^{6} h(x_i) = 5 + 8 + 14 + 16 + 5 + 2 = 50$$

Durchschnittsnote:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{j} x_i * h(x_i) = \frac{1}{50} * (1 * 5 + 2 * 8 + 3 * 14 + 4 * 16 + 5 * 5 + 6 * 2) = \frac{164}{50} = 3,3$$

ARITHMETISCHES MITTEL AUS HÄUFIGKEITSTABELLEN

- **Arithmetisches Mittel** $\overline{x}$

Fall 2: Relative Häufigkeit $f(x_i) = \frac{h(x_i)}{n}$

$$\overline{x} = \sum_{i=1}^{j} x_i * f(x_i) = (x_1 f(x_1) + x_2 f(x_2) + \cdots + x_j f(x_j))$$

$f(x_i)$ *relative Häufigkeit der Merkmalsausprägung* x_i

n *Summe der absoluten Häufigkeiten*

j *Anzahl der Merkmalsausprägungen* x_i

ARITHMETISCHES MITTEL AUS HÄUFIGKEITSTABELLEN

- **Arithmetisches Mittel** $\bar{x}$

Beispiel:

Berechnung des arithmetischen Mittels über die relativen Häufigkeiten:

Note x_i	1	2	3	4	5	6
Anzahl Schüler $h(x_i)$	5	8	14	16	5	2
Relative Häufigkeit $f(x_i) = h(x_i)/\mathrm{n}$	0,1	0,16	0,28	0,32	0,1	0,04

Schüler insgesamt:

$$n = \sum_{1}^{6} h(x_i) = 5 + 8 + 14 + 16 + 5 + 2 = 50$$

Durchschnittsnote:

$$\bar{x} = \sum_{i=1}^{j} x_i * f(x_i) = 1 * 0,1 + 2 * 0,16 + 3 * 0,28 + 4 * 0,32 + 5 * 0,1 + 6 * 0,04 = 3,3$$

ARITHMETISCHES MITTEL BEI KLASSIERTEN DATEN

Näherungswert für Arithmetisches Mittel

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} m_i * h_i = \sum_{i=1}^{k} m_i * f_i$$

$m_1, \ldots, m_k$	***Klassenmitten (!!)***
$h_1, \ldots, h_k$	***absolute Klassenhäufigkeiten***
$f_1, \ldots, f_k$	***relative Klassenhäufigkeiten***

→ Näherungswert, weil die Verteilung in den Klassen nicht bekannt ist. Es wird angenommen, dass die Beobachtungswerte jeweils in den Klassenmitten liegen.

ARITHMETISCHES MITTEL BEI KLASSIERTEN DATEN

- **Arithmetisches Mittel** $\overline{x}$

Fall 1: Absolute Häufigkeit h_i

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} m_i * h_i = \frac{1}{n} * (m_1 h_1 + m_2 h_2 + \cdots + m_k h_k)$$

$$n = \sum_{i=1}^{k} h_i = h_1 + h_2 + \cdots + h_k$$

h_i	*absolute Häufigkeit der i-ten Klasse*
n	*Summe der absoluten Häufigkeiten*
k	*Anzahl der Klassen*
m_i	*Klassenmitte der i-ten Klasse*

ARITHMETISCHES MITTEL BEI KLASSIERTEN DATEN

- **Arithmetisches Mittel** $\bar{x}$

Beispiel:

klassierte Häufigkeitstabelle für das Körpergewicht, Berechnung über die <u>absoluten</u> Häufigkeiten:

Klasse x_i	41 bis 50	51 bis 60	61 bis 70	71 bis 80	81 bis 90
Häufigkeit h_i	20	15	10	4	1
Klassenmitte m_i	45,5	55,5	65,5	75,5	85,5

Der Häufigkeit wird die Klassenmitte zugeordnet. Man unterstellt, dass alle 15 Schüler z. B. der Klasse x_2 das Körpergewicht 55,5 kg haben.

Schüler insgesamt:

$$n = \sum_{1}^{5} h_i = 20 + 15 + 10 + 4 + 1 = 50$$

Durchschnittsgewicht**:**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{5} h_i * m_i = \frac{1}{50} * (20 * 45,5 + 15 * 55,5 + 10 * 65,5 + 4 * 75,5 + 1 * 85,5) = \frac{2785}{50} = 55,7$$

ARITHMETISCHES MITTEL BEI KLASSIERTEN DATEN

- **Arithmetisches Mittel** $\overline{x}$

Fall 2: Relative Häufigkeit $f_i = \frac{h_i}{n}$

$$\overline{x} = \sum_{i=1}^{k} m_i * f_i = m_1 f_1 + m_2 f_2 + \cdots + m_k f_k$$

f_i	*relative Häufigkeit der i-ten Klasse*
n	*Summe der absoluten Häufigkeiten*
k	*Anzahl der Klassen*
m_i	*Klassenmitte der i-ten Klasse*

ARITHMETISCHES MITTEL BEI KLASSIERTEN DATEN

- **Arithmetisches Mittel** $\bar{x}$

Beispiel:

klassierte Häufigkeitstabelle für das Körpergewicht, Berechnung über die relativen Häufigkeiten:

Klasse x_i	41 bis 50	51 bis 60	61 bis 70	71 bis 80	81 bis 90
Häufigkeit h_i	20	15	10	4	1
Relative Häufigkeit $f_i = h_i/n$	0,40	0,30	0,20	0,08	0,02
Klassenmitte m_i	45,5	55,5	65,5	75,5	85,5

Schüler insgesamt:

$$n = \sum_{1}^{5} h_i = 20 + 15 + 10 + 4 + 1 = 50$$

Durchschnittsgewicht:

$$\bar{x} = \sum_{i=1}^{5} m_i * f_i = 45,5 * 0,40 + 55,5 * 0,30 + 65,5 * 0,20 + 75,5 * 0,08 + 85,5 * 0,02 = 55,7$$

MEDIAN VS. MITTELWERT

Beispiel:

9 Personen aus einer Abteilung haben folgende Einkünfte in €:

1.200	1.050	950	1.100	900	1.800	6.600	1.150	1.000

$$\boldsymbol{\bar{x} = 1.660}$$

Dieser Durchschnitt liefert ein falsches Bild, weil die Mehrzahl (7 von 9 Personen) höchstens 1.200 € verdient. Der Wert 6.600 € zieht den Mittelwert nach oben. Man sucht nach einem Wert, der die Verteilung der Einkünfte besser charakterisiert. Dazu werden die Verdienste der Größe nach sortiert:

900	950	1.000	1.050	1.100	1.150	1.200	1.800	6.600

$$\bar{x}_z = x_{\frac{n+1}{2}} = x_{\frac{9+1}{2}} = x_5 = 1.100$$

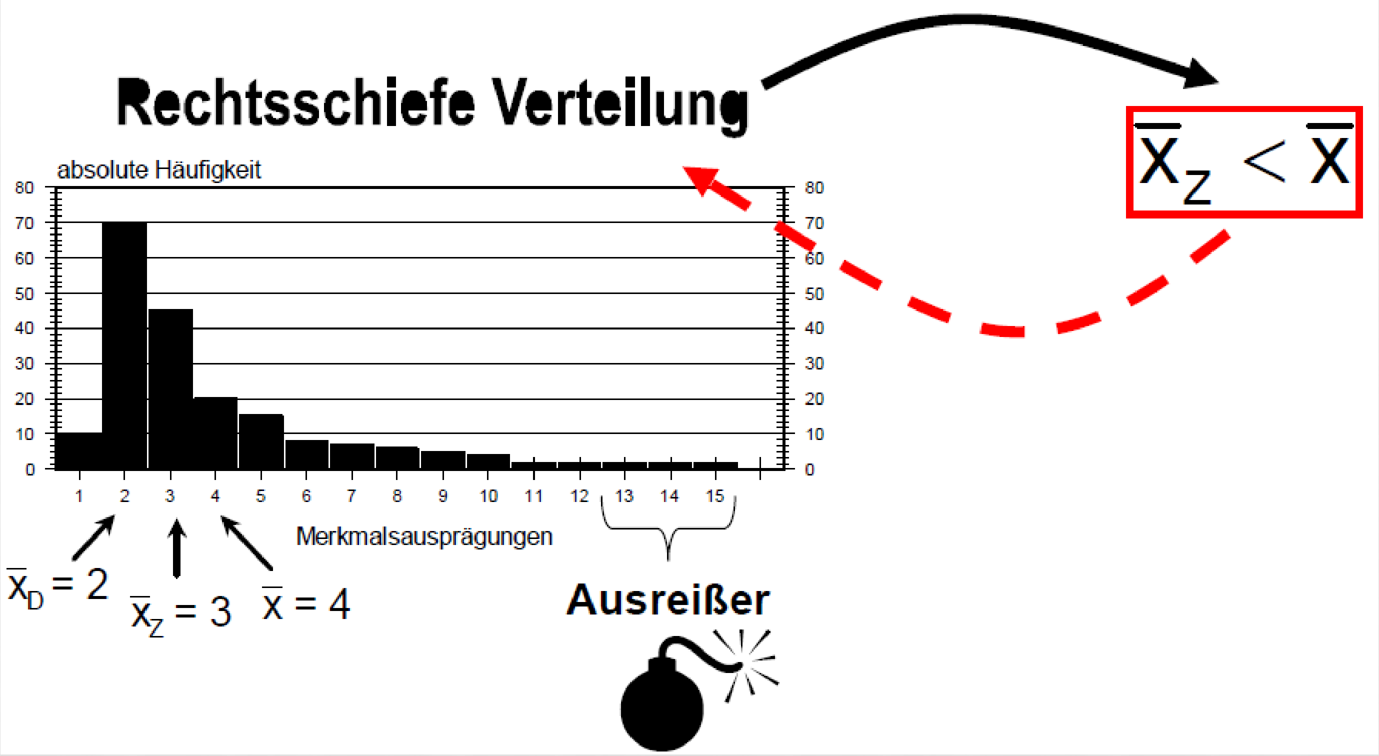
Der Median beschreibt die Verteilung besser als der Mittelwert, Ausreißer haben auf den Median keinen Einfluss.

NEIGUNG / SCHIEFE

Folgende Faustregel setzt Modus, Median und arithmetisches Mittel in Beziehung:

rechtsschiefe (linkssteile) Häufigkeitsverteilung:

Modus < Median < arithmetisches Mittel $\bar{x}_D < \bar{x}_Z < \bar{x}$

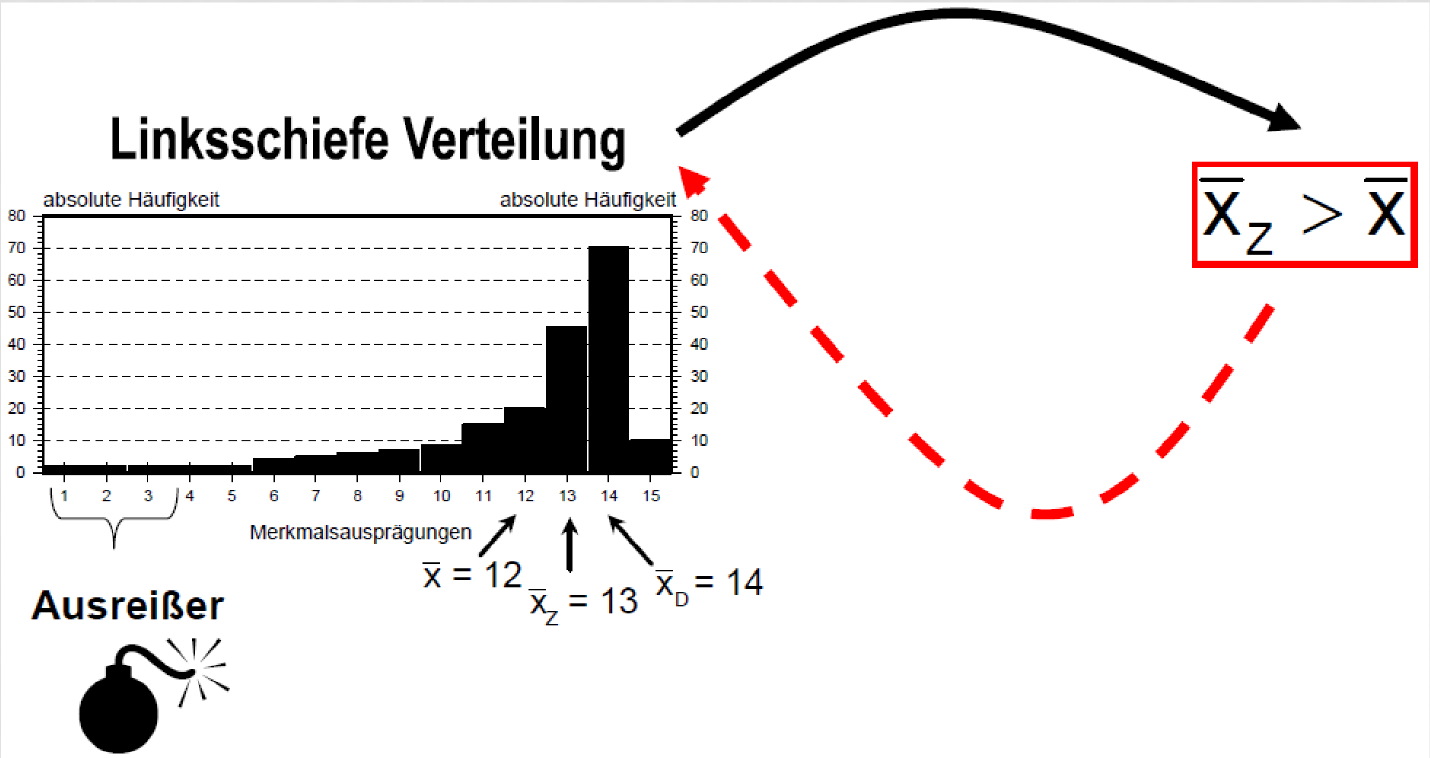


NEIGUNG / SCHIEFE

Folgende Faustregel setzt Modus, Median und arithmetisches Mittel in Beziehung:

linksschiefe (rechtssteile) Häufigkeitsverteilung:

Modus > Median > arithmetisches Mittel $\bar{x}_D > \bar{x}_Z > \bar{x}$

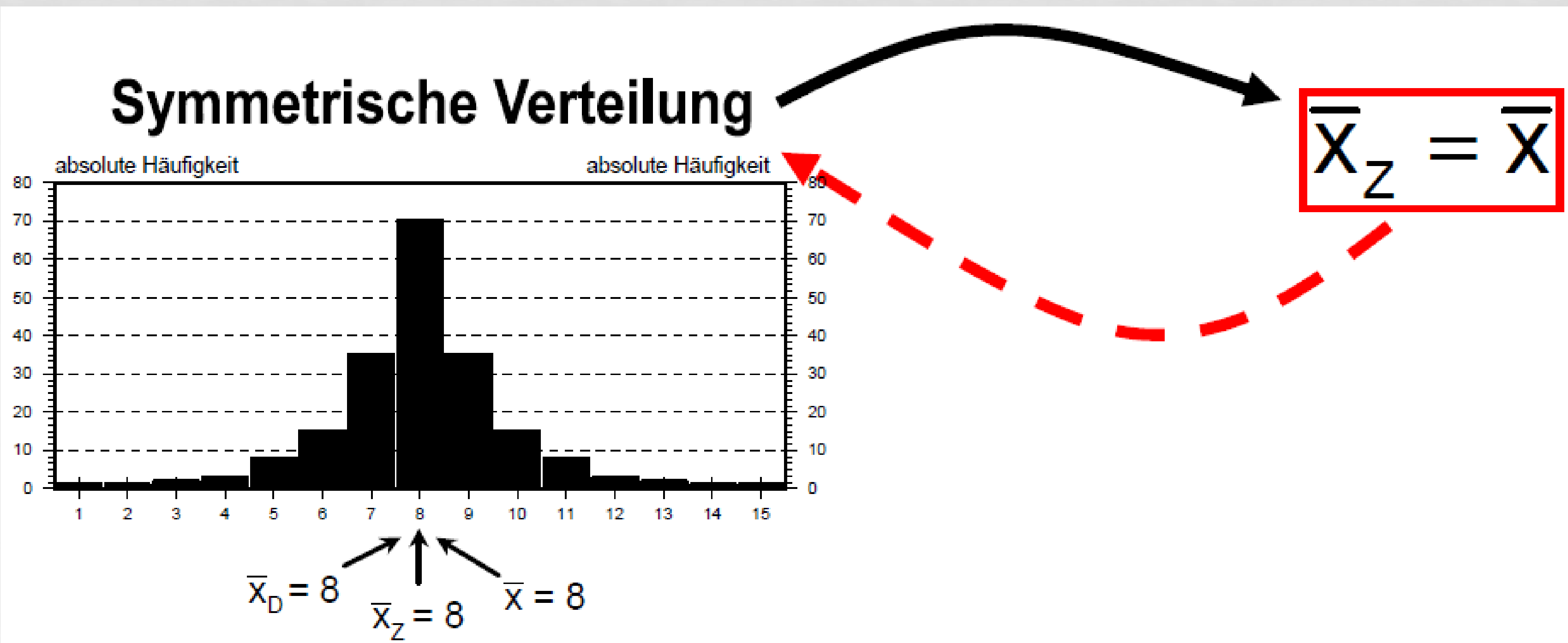


NEIGUNG / SCHIEFE

Folgende Faustregel setzt Modus, Median und arithmetisches Mittel in Beziehung:

unimodale symmetrische Häufigkeitsverteilung:

Modus ≈ Median ≈ arithmetisches Mittel $\bar{x}_D \approx \bar{x}_Z \approx \bar{x}$



BEISPIELÜBUNG

Modus, Median, arithmetisches Mittel, Schiefe

x_i	$h(x_i)$	$f(x_i)$ (%)	$H(x_i)$	$F(x_i)$ (%)
1	1	8,3%	1	8,3%
2	6	50,0%	7	58,3%
3	2	16,7%	9	75,0%
4	2	16,7%	11	91,7%
9	1	8,3%	12	100,0%
Summe	**12**	**100,0%**	**-**	**-**

$$\bar{x}_D = 2 \qquad \bar{x}_Z = 2$$

$$\bar{x} = \frac{1}{12}(1 \cdot 1 + 2 \cdot 6 + 3 \cdot 2 + 4 \cdot 2 + 9 \cdot 1) = \frac{36}{12} = 3$$

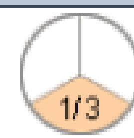
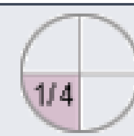
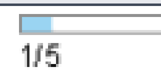
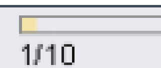
wegen $\bar{x}_Z < \bar{x}$ rechtsschiefe Verteilung

QUANTILE

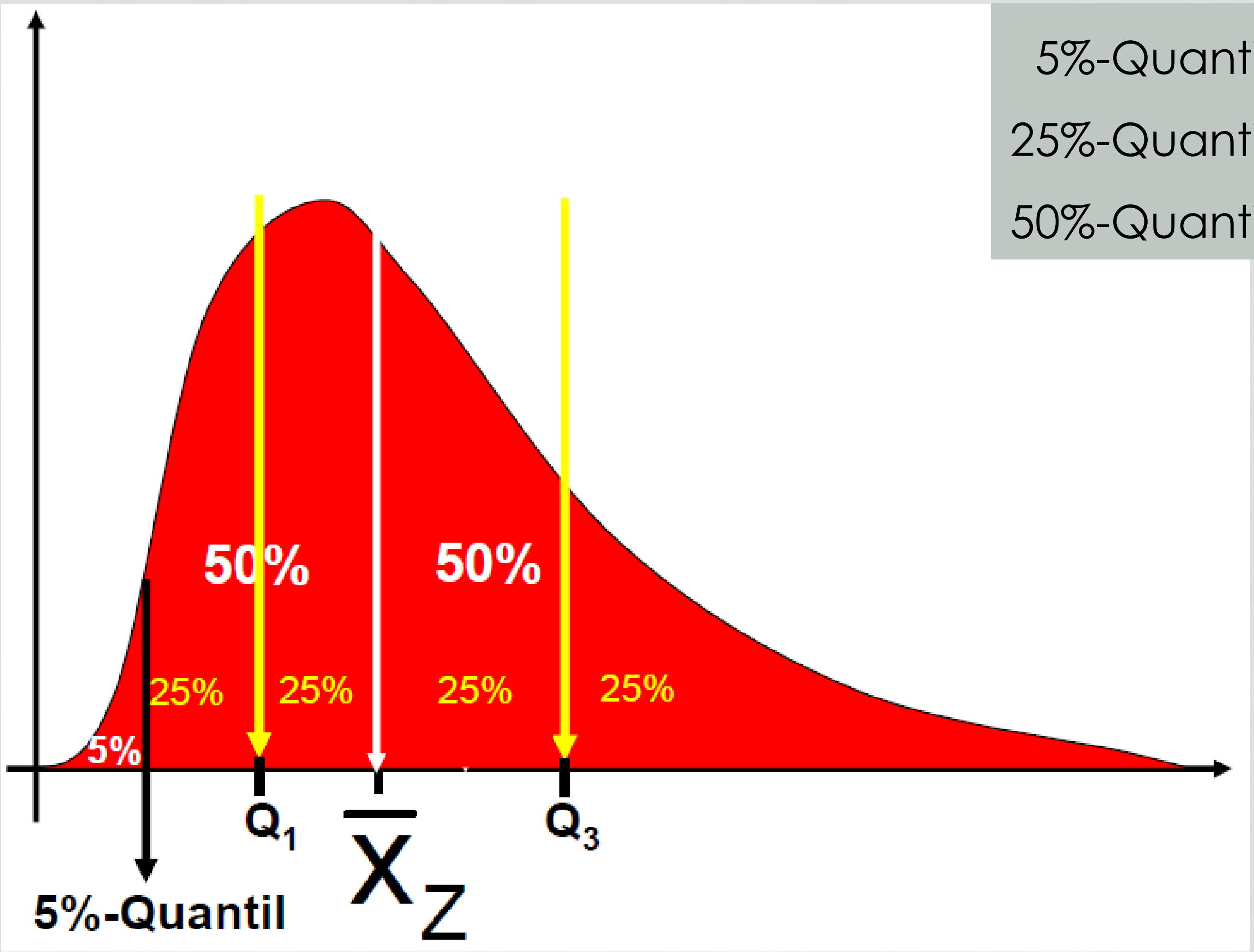
Ein **Quantil** ist ein Lagemaß in der Statistik.

Quantile teilen eine Verteilung in Abschnitte gleicher Häufigkeit.

Spezielle Quantile:

Benennung der Quantile $\tilde{x}_p$	Anzahl der Intervalle	
Terzile	3	
Quartile	4	
Quintile	5	
Dezile	10	
Vigintile	20	
Perzentile (Zentile)	100	

QUANTILE



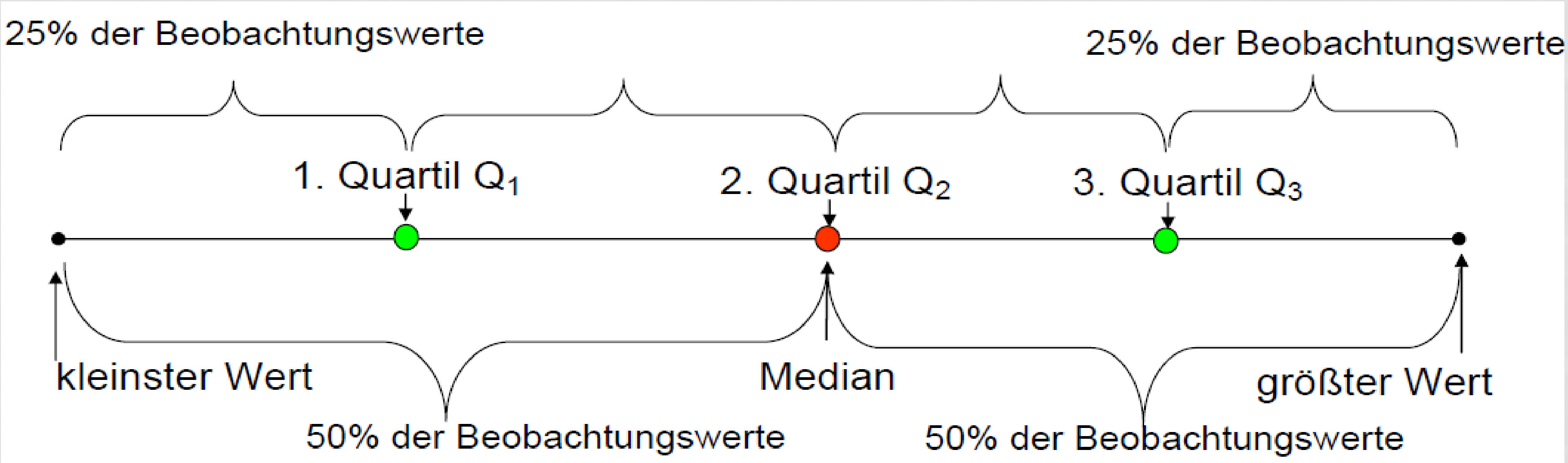
5%-Quantil = Quintil

25%-Quantil = Quartil = Q1, Q3

50%-Quantil = Median = Q2

QUARTILE

5-Punkte-Zusammenfassung der geordneten statistischen Reihe:



Der **Median** teilt einen nach Größe sortierten Datensatz in der Mitte

→ links und rechts vom Median liegen gleich viele Beobachtungswerte. Unterteilt man die linke und die rechte Hälfte nach gleicher Vorschrift, wie man den Median bestimmt, so erhält man 4 gleich große Bereiche, die durch drei Quartils aufgeteilt werden.

25% aller geordneten Beobachtungswerte sind kleiner als das **1. Quartil**.

50% aller geordneten Beobachtungswerte sind kleiner als das **2. Quartil**.

75% aller geordneten Beobachtungswerte sind kleiner als das **3. Quartil**.