Maße der zentralen Tendenz (10)

- Die Berechnung der zentralen Tendenz bei kategorisierten Daten mit offenen Endklassen I -

Beispiel 1: offene Endklasse

Alter	k	x_k	f_k	$x_k \cdot f_k$	p_k	$x_k \cdot p_k$	$p_{cum\,k}$
bis 20	1	?	3	?	0.06	?	0.06
21 - 25	2	23	20	460	0.40	9.20	0.46
26 - 30	3	28	17	476	0.34	9.52	0.80
31 u. älter	4	?	10	?	0.20	?	1.00
Σ			N=50	?	1.00	AM= ?	

Modus: 23

Median: 26.1 berechenbar nach

$$C_{50} = IA_k + b \cdot \frac{0.50 - p_{cum}(IA_k)}{p_k}$$

AM: nicht berechenbar nach

$$AM = \frac{\sum_{k=1}^{m} x_k \cdot f_k}{N} = \sum_{k=1}^{m} x_k \cdot p_k$$

Maße der zentralen Tendenz (11)

- Die Berechnung der zentralen Tendenz bei kategorisierten Daten mit offenen Endklassen II -

Beispiel 2: Median/Modus fällt in offene Endklasse

Alter	k	x_k	f_k	p_k	$p_{cum\,k}$
bis 20	1	?	3	0.06	0.06
21 - 25	2	23	7	0.14	0.20
26 - 30	3	28	10	0.20	0.40
31 u. älter	4	?	30	0.60	1.00
Σ			N=50	1.00	

Modus: nicht angebbar

Median: nicht berechenbar

AM: nicht berechenbar

Maße der zentralen Tendenz (12)

- Die Minimumseigenschaften der drei Maße zentraler Tendenz I -

Wenn man einen Kennwert der zentralen Tendenz angibt, dann möchte man, dass er die Gesamtheit der beobachteten Messwerte möglichst gut repräsentiert.
D.h. das Maß der zentralen Tendenz soll „irgendwie insgesamt“ minimal von den beobachteten Messwerten abweichen.
Modus, Median und arithmetisches Mittel weisen eine solche Minimumseigenschaft des „ irgendwie insgesamt minimal“ in verschiedener Weise auf.

Minimumseigenschaft des Modus

Die Wahrscheinlichkeit p_k, dass eine beliebige Person der Stichprobe aus der Modalkategorie kommt, ist maximal, weil der Modus der relativ häufigste Wert ist.
D.h. die Wahrscheinlichkeit des Irrtums (1- p_k), wenn ich bei einer beliebigen Person das Vorliegen des Modalwerts annehme, ist verglichen mit allen anderen Werten minimal.

Maße der zentralen Tendenz (13)

- Die Minimumseigenschaften der drei Maße zentraler Tendenz II -

Minimumseigenschaft des Median

Beim Median ist die Summe der absoluten Abweichungen aller Messwerte minimal im Vergleich zu allen anderen denkbaren Skalenwerten.
Formal:

$$\sum_{i=1}^{N} |x_i - a| = \min \Leftrightarrow a = Md(x)$$

Wird der Median als Schätzwert für einen zufällig gezogenen Messwert genommen, ist der durchschnittliche absolute Schätzfehler minimal.

Maße der zentralen Tendenz (14)

- Die Minimumseigenschaften der drei Maße zentraler Tendenz III -

Minimumseigenschaft des AM

Um ein Maß der zentralen Tendenz zu erhalten, dass große Abweichungen stärker berücksichtigt als kleine, kann man die Abweichungen quadrieren. Beim arithmetischen Mittel ist die Summe der quadrierten Abweichungen (kurz: Quadratsumme QS) minimal.

Formal:

$$\sum_{i=1}^{N}(x_i - a)^2 = \min \Leftrightarrow a = \bar{x}$$

Wird das arithmetische Mittel als Schätzwert für einen zufällig gezogenen Messwert genommen, ist der durchschnittliche ‚quadrierte Schätzfehler' minimal.

Maße der zentralen Tendenz (15)

- Die Minimumseigenschaften der drei Maße zentraler Tendenz IV -

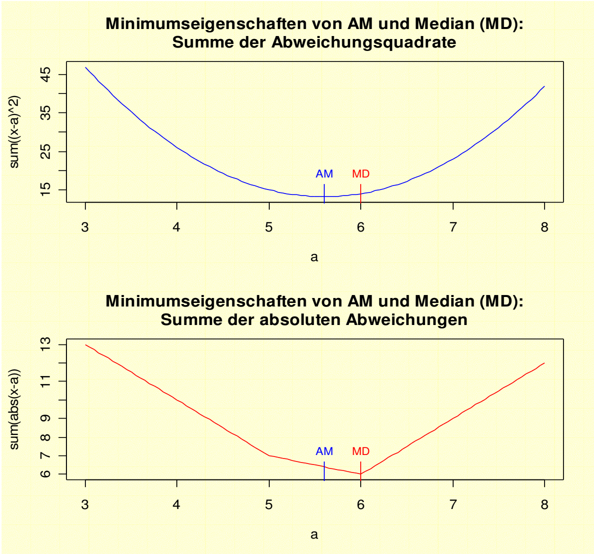
Beispiel Wirtz & Nachtigall (1998), S. 75f.:

x_i	$x_i - 5.6$	$x_i - 6.0$	$\lvert x_i - 5.6\rvert$	$\lvert x_i - 6.0\rvert$	$(x_i - 5.6)^2$	$(x_i - 6.0)^2$
3	-2.6	-3.0	2.6	3.0	6.76	9.0
6	0.4	0.0	0.4	0.0	0.16	0.0
8	2.4	2.0	2.4	2.0	5.76	4.0
6	0.4	0.0	0.4	0.0	0.16	0.0
5	-0.6	-1.0	0.6	1.0	0.36	1.0
$\sum$	**0.0**	**-2.0**	**6.4**	**6.0**	**13.2**	**14.0**

AM = (3+6+8+6+5)/5 = **5.6** Median = **6.0** Modus = 6.0

Maße der zentralen Tendenz (16)

- Die Minimumseigenschaften der drei Maße zentraler Tendenz V -



Maße der zentralen Tendenz (17)

Eigenschaften der Maße zentraler Tendenz
- Zusammenfassender Vergleich I -

	Modus	*Median*	*arithmetisches Mittel*
verbale Definition	häufigster Wert einer Verteilung	mittlerer Wert einer Verteilung	"Schwerpunkt" einer Verteilung
Mindestskalenniveau zur sinnvollen Interpretation	nominal	ordinal	intervall
Transformationen, die mitgemacht werden	eineindeutig	monoton	linear
Minimumseigenschaft	1-relative Häufigkeit der Modalklasse (="Vorhersage-Irrtum für eine beliebige Person der Stichprobe")	Summe der Abweichungsbeträge	Summe der Abweichungs-quadrate (Das AM ist der optimale Schätzwert nach dem "Kriterium der kleinsten Quadrate")
Summe der mit Vorzeichen versehenen Abweichungen der Messwerte	unbestimmt	unbestimmt	immer gleich 0

Maße der zentralen Tendenz (18)

Eigenschaften der Maße zentraler Tendenz
- Zusammenfassender Vergleich II -

	Modus	*Median*	*arithmetisches Mittel*
Bestimmung aus gruppierten Daten	Wert der Kategorienmitte der Modalklasse	Interpolation des 50. Centils (nur bei künstlich diskreten Daten)	Summe über das Produkt aus Kategorienmitte • relativer Häufigkeit
Bestimmung aus gruppierten Daten mit offenen Endklassen	möglich, wenn Modalklasse nicht offene Endklasse	möglich, wenn 50. Centil nicht in offene Endklasse fällt	nicht möglich
Beziehung zur Darstellung als Säulendiagramm (Modus) bzw. Histogramm (Median, AM)	Kategorie mit der höchsten Säule	Die Senkrechte auf den Median halbiert die Fläche unter dem Histogramm	Der Schwerpunkt des Histogramms befindet sich genau über dem AM
Reaktion auf Ausreißerwerte	nicht relevant	unsensibel	sensibel, besonders in kleinen Stichproben
relative Lage in Abhängigkeit von der Schiefe der Verteilung	linkssteile Verteilung: AM > Md > Mod rechtssteile Verteilung: Mod > Md > AM symmetrische Verteilung: AM = Md = Mod		

Maße der zentralen Tendenz (19)

- Der gewichtete Durchschnitt I -

Bei der Berechnung des gewichteten Durchschnitts erhalten die Messwerte bei der Summenbildung ein verschiedenes Gewicht und gehen nicht - wie bei der regulären Durchschnittsbildung - mit gleichem Gewicht ein.

$$\bar{X}_{gewichtet} = \frac{\sum_i x_i \cdot g_i}{\sum_i g_i}$$

mit:
x_i = i-ter Messwert
g_i = Gewicht, das dem i-ten Messwert zukommt

$$GAM = \frac{\sum_{j=1}^{k} n_j \cdot \bar{x}_j}{\sum_{j=1}^{k} n_j}$$

GAM zur Berechnung eines Gesamtmittelwertes aus verschiedenen Einzelmittelwerten

mit:
k = Anzahl der Kollektive
n_j = Größe des Kollektivs j
$\bar{x}_j$= AM der Kollektivs j

Maße der zentralen Tendenz (20)

- Der gewichtete Durchschnitt II -

Beispiel: Berechnung einer gewichteten Durchschnittsnote

Fach	*Note* x_i	*Gewicht* g_i	$x_i \cdot g_i$
Deutsch	2	2	4
Mathematik	3	2	6
Englisch	1	2	2
Physik	2	1	2
Biologie	1	2	2
Chemie	2	1	2
Geschichte	4	2	8
Erdkunde	3	0.5	1.5
Musik	4	0.5	2
Sport	4	0	0
Kunst	3	1	3
Religion	3	0	0
Summe	32	14	32.50

ungewichtet:

$$\bar{X} = \frac{\sum_i x_i}{N} = \frac{32}{12} \cong 2.67$$

gewichtet:

$$\bar{X}_{gewichtet} = \frac{\sum_i x_i \cdot g_i}{\sum_i g_i} = \frac{32.5}{14} \cong 2.32$$