

## Statistische Kennwerte

Sagen uns tabellarische und graphische Darstellungen etwas über die Verteilung der *einzelnen* Werte einer Stichprobe, so handelt es sich bei *statistischen Kennwerten* um eine Kennzahl, die - stellvertretend für die vielen einzelnen Werte - pauschal über die Gesamtverteilung einer Variable Auskunft gibt. Derartige Zahlen nennt man Maße der zentralen Tendenz. Nun gibt es verschiedene derartige Maßzahlen. Für *welches* Maß Sie sich entscheiden, hängt unter anderem vom *Skalenniveau* Ihrer Daten ab. Zusätzlich zu den Maßen der zentralen Tendenz zählen auch die so genannten Dispersions- oder Streuungsmaße zu den statistischen Kennwerten. Wir besprechen sie anschließend an die Maße der zentralen Tendenz.

### **Maße der zentralen Tendenz**

Die drei am häufigsten verwendeten sind:

1) der Mittelwert, auch das arithmetische Mittel genannt (der durchschnittliche Werte)

Voraussetzung für das arithmetische Mittel ist **Intervallskalenniveau**.

2) der Median (der Wert, der die der Größe nach geordneten Werte in genau zwei gleich große Hälften teilt)

Voraussetzung für den Median ist **Ordinalskalenniveau**

3) der Modalwert (der am häufigsten vorkommende Wert)

Voraussetzung für den Modalwert ist **Nominalskalenniveau**.

Der Modalwert stellt nach dieser Aufstellung die geringsten Anforderungen an das Skalenniveau (Nominalskala), das arithmetische Mittel die höchsten Anforderungen (Intervallskala).

Darüber hinaus repräsentieren alle drei genannten Maße der zentralen Tendenz - Mittelwert, Median und Modalwert - auf *verschiedene* Weise die Gesamtverteilung der Ausprägungen einer Variablen. Auf welche Weise, zeigen nachstehende Ausführungen.

## 1) das arithmetische Mittel:

Voraussetzung für die Berechnung des arithmetischen Mittels ist - wie bereits erwähnt - Intervallskalenniveau. Dieses Skalenniveau ist in den Sozialwissenschaften eher selten anzutreffen - wie in vielen Lehrbüchern auch freimütig eingestanden wird. Dennoch gehen die meistens inferenzstatistischen Methoden, die dann in diesen Lehrbüchern dargestellt werden, vom Mittelwert aus.

Zur Berechnung des Mittelwerts wird eine bestimmte Notation verwendet, das Summenzeichen, das ich denn nun kurz erläutern möchte.

*Zur Notation des Summenzeichens*

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

Man liest dies wie folgt: Summe von  $x_i$ , wobei  $i$  von 1 bis  $n$  geht.  $n$  ist die Gesamtanzahl aller Werte, für die wir die Summe berechnen wollen,  $x_i$  bezeichnet die einzelnen Werte. Stellen wir uns vor, wir müssten die Summe von den drei Werten 5, 7, 10 berechnen. Dann ist  $x_1 = 5$ ;  $x_2 = 7$ ;  $x_3 = 10$ .  $x$  ist also einfach irgendeine Variable und die Indizes hinter der Variable (Beispiel:  $x_1$ ) kennzeichnen die einzelnen Werte dieser Variablen.  $i$  bezeichnet man auch als *Laufvariable*. Je nachdem, welche Zahlen man in die Werte dieser Variablen einsetzt, bekommt man verschiedene Werte der Variablen  $x$ . Setzen wir die Zahl 1 ein, so bekommen wir den ersten Wert, die Zahl 2 den zweiten Wert usw.

Das Summenzeichen ist nur eine abgekürzte Schreibweise für die Summenbildung der drei Werte  $x_1 + x_2 + x_3$ . Bei nur drei Werten spielt diese Abkürzung keine Rolle, wichtig wird sie bei vielen Werten. Statt bei 30 Versuchspersonen alle einzelnen Werte zur Summenbildung anführen zu müssen ( $x_1 + x_2 + x_3 + \dots x_{30}$ ), schreiben wird also:

$$\sum_{i=1}^{30} x_i$$

Das arithmetische Mittel erhalten wir nun, indem wir die Summe durch die Gesamtanzahl der summierten Werte dividieren, im vorliegenden konkreten Falle also

$$\frac{\sum_{i=1}^{30} x_i}{30}$$

Dieser so gewonnene Mittelwert ist nun eine Kennzahl, die auf eine bestimmte Weise die 30 einzelnen Werte repräsentiert. Wir können diese Kennzahl auch als eine *Schätzung* der einzelnen Werte interpretieren.

Nun hat das arithmetische Mittel ganz bestimmte Eigenschaften: Am besten veranschaulicht man sich den Mittelwert, indem man ihn mit dem so genannten *Massenmittelpunkt*, auch *Schwerpunkt* genannt, vergleicht.

Eine gängige Definition des Schwerpunktes lautet: Ein im Schwerpunkt unterstützter Körper ist bei alleiniger Wirkung der Schwerkraft in jeder beliebigen Lage im stabilen Gleichgewicht.

Dazu folgendes Beispiel: Stellen Sie sich die Reaktionszeiten gemessen in Millisekunden angeordnet auf einer Skala vor. Die Skala reiche von 9 ms - dem kleinsten Wert - bis zu 31 ms - dem größten Wert. 15 ms kommen beispielsweise dreimal vor. Stellen Sie sich weiters vor, die Skala wäre eine Balancestange und jeder der eingetragenen Werte hätte das gleiche Gewicht. Wenn nun die Stange an einem ganz bestimmten Punkt unterstützt wird, welchen Punkt auf der Balancestange müssten Sie finden, damit sie sich im Gleichgewicht befindet?

Dieser Punkt auf der Stange entspricht dem Mittelwert. Werte, die vom Mittelwert weiter entfernt sind, haben ein stärkeres Gewicht. Ebenfalls wirkt sich auf das Gewicht aus, *wie oft* ein Wert vorkommt (da 15 in unserem Beispiel 3-mal vorkommt, verstärkt sich das Gewicht von 15)

Das bedeutet: *Der Mittelwert ist jener Wert, der Unterschiede in der einen Richtung und Unterschiede in der anderen Richtung genau ausbalanciert.*

Mathematisch gerechnet bedeutet dies:

Die Summe der Abweichungen der einzelnen Werte vom Mittelwert unter Berücksichtigung des Vorzeichens ist Null.

Beispiel (der Mittelwert folgender Werte ist 22.4333):

Meßwert	Differenz zum Mittelwert
15	-7.4333
20	-2.4333
17	-5.4333
35	+12.566
19	-3.4333
22	-0.4333
24	+1.566
33	+10.566
26	+3.566
15	-7.4333
41	+18.566
22	-0.4333
11	-11.4333
21	-1.4333
15	-7.4333
25	+2.566
24	+1.566
21	-1.4333
31	+8.566
27	+4.566
31	+8.566
9	-13.4333
23	+0.566
21	-1.4333
25	+2.566
28	+5.566
17	-5.4333
13	-9.4333
24	+1.566
18	-4.4333

Zählen Sie die Differenzen in der rechten Spalte zusammen, so werden Sie – bis auf Rundungsfehler – die Zahl 0 erhalten.

Stellen Sie sich nun vor, Sie würden rein zufällig aus den 30 möglichen Werten einen Wert ziehen. Jedes Mal würden Sie nicht *genau* den Mittelwert ziehen, sondern einen Wert, der nach unten oder oben abweicht. Das Ausmaß des Irrtums bei jeder Ziehung ist  $d_i$ .

Nun hat der Mittelwert die Eigenschaft, dass die *Summe der möglichen Fehler bei allen Ziehungen Null ist*.

### ***Veränderungen des Mittelwertes bei Lineartransformationen***

Lineartransformationen haben allgemein den folgenden Charakter:

$y = b \cdot x + a$ , wobei  $b$  ein multiplikativer Faktor ist (der auch 1 sein kann) und  $a$  eine Konstante (die auch 0 sein kann).

Beispiele:

$$y = 2 \cdot x + 3$$

$$y = x + 2$$

$$y = 3 \cdot x$$

Nun gilt: *Der Mittelwert der linear transformierten Werte ist gleich dem linear transformierten Mittelwert der Werte.*

Das bedeutet: Führen wir eine Lineartransformation für alle Werte durch (Beispiel:  $2 \cdot x + 3$ ) und berechnen dann anschließend für die lineartransformierten Werte den neuen Mittelwert, so kommen wir zu dem gleichen Ergebnis, wenn wir auf den ursprünglichen Mittelwert der noch nicht transformierten Werte die gleiche Lineartransformation anwenden.

Bezogen auf die oben aufgelisteten Werte mit dem Mittelwert 22.4333 bedeutet dies für die Lineartransformation  $y = 2 \cdot x + 3$ : Der Mittelwert der lineartransformierten Werte beträgt:  $2 \cdot 22.4333 + 3 = 47.8666$ , was man leicht nachprüfen kann.

## 2) der Median

Der Median ist jener Wert, der die der *Größe nach geordneten Werte* in genau zwei Hälften teilt. Damit setzt der Median nur *Ordinalskalenniveau* voraus. Denn auch ordinalskalierte Daten sind ja der Größe nach geordnet - nur der *Abstand* zwischen den Werten spielt keine Rolle.

Der Median hat die folgende Eigenschaft:

Ziehen Sie per Zufall aus  $n$  Werten einen beliebigen Wert, so ist die Chance, einen Wert kleiner als den Median zu ziehen gleich groß wie die Chance einen Wert größer als den Median zu ziehen. Der Grund dafür liegt einfach daran, dass unter dem Median genauso viele Fälle liegen als über dem Median.

Zur *Berechnung* des Median:

### a) **Intervallskalierte Daten mit kleiner Fallanzahl**

Wir unterscheiden zwei Situationen:

*Die Gesamtanzahl aller Werte ist ungerade:*

- Sie ordnen alle Werte der Größe nach (von links nach rechts)

- Sie addieren zur Gesamtzahl + 1 und dividieren die erhaltene Summe durch 2. Dadurch erhalten Sie eine Zahl, nennen wir sie Z.
- Sie zählen von links nach rechts z-mal und erhalten so den Median.

#### Beispiel:

Unsere Werte sind: 3, 5, 7, 8, 10. Sie wurden der Einfachheit halber bereits der Größe nach geordnet.

Insgesamt haben wir 5 Werte, also eine ungerade Anzahl. Zu der Zahl 5 zählen wir +1 und erhalten die Zahl 6. Diese Zahl dividieren wir durch 2 und erhalten die Zahl 3.

Der Median ist also die 3-te Zahl von links gezählt, also die Zahl 7.

Beachten Sie bitte, dass Sie *vorher* alle Werte der Größe nach ordnen müssen.

*Die Gesamtanzahl aller Werte ist gerade:*

- Sie ordnen wiederum alle Werte der Größe nach von links nach rechts.
- Sie teilen die Gesamtanzahl aller Werte in zwei Hälften.
- der Median wird berechnet, indem Sie den Mittelwert zwischen dem größten Wert der ersten Hälfte und dem kleinsten Wert der zweiten Hälfte bilden.

#### Beispiel:

Unsere Werte sind: 3, 5, 7, 8. Auch diese Werte wurden bereits der Größe nach geordnet.

Die insgesamt 4 Werte bilden die beiden Hälften "3, 5" und "7, 8".

Die Zahl "5" ist der größte Wert der ersten Hälfte und die Zahl "7" der kleinste Wert der zweiten Hälfte.

$7+5$  ergibt 12. 12 dividiert durch 2 ergibt 6.

"6" ist also der gewünschte Median. Auch hier gilt es zu beachten, dass die Werte *vorher* der Größe nach geordnet werden müssen.

#### **b) Intervallskalierte Daten mit großer Fallanzahl**

Es handelt sich hierbei in der Regel um *gruppierte* (also in Kategorien zusammengefasste Werte). Bei gruppierten Werten ist der Median jener Wert, unterhalb dessen 50 % aller Fälle liegen.

Man kann ihn grafisch über das Summenpolygon bestimmen. Beim Summenpolygon wird die y-Achse mit den kumulierten Prozentsätzen beschriftet. Man zieht in einem ersten Schritt bei 50 % eine Parallele zur x-Achse. An der Stelle, an der diese Parallele das Summenpolygon schneidet, fällt man ein senkrechtes Lot zur x-Achse. Die Stelle, an der dieses senkrechte Lot die x-Achse schneidet, ist der gesuchte Median. Dieser ist im vorliegenden Beispiel 22.5.

### c) Ordinalskalierte Daten

Fall 1) Ungerade Anzahl:

Gehen wir von dem bereits bekannten Beispiel aus. Unsere Werte sind:

3, 5, 7, 8, 10

In diesem Falle wird der Median gleich wie bei Intervallskalenniveau berechnet - es handelt sich wiederum um den Wert in der Mitte, also die Zahl 7!

Fall 2) gerade Anzahl:

Beispiel: 3, 5, 7, 8

In *diesem* Falle ergibt es keinerlei Sinn, den Mittelwert von 5 und 7 zu berechnen. Denn bei ordinalskalierten Daten ist das Intervall zwischen 5 und 7 überhaupt nicht metrisch definiert!

Bei ordinalskalierten Daten ist der Median - streng genommen - das Intervall zwischen 5 und 7!

### 3) der Modalwert

Dies ist einfach der am häufigsten vorkommende Wert. Der Modalwert ist jener Wert, dessen Wahrscheinlichkeit rein zufällig aus den 30 Werten gezogen zu werden, am größten ist. Bei gruppierten Daten ist dies jene Klasse, die am häufigsten besetzt ist.

Voraussetzung zur Berechnung des Modalwertes ist *Nominalskalenniveau*.

Abschließend ist zu den drei genannten Maßen der zentralen Tendenz - Mittelwert, Median, Modalwert - folgendes zu sagen:

1) Jedes der Maße kann nur unter der Voraussetzung des für es geltende Skalenniveau berechnet werden. (Mittelwert: Intervallskala; Median: Ordinalskala; Modalwert: Nominalskala)

Umgekehrt kann man allerdings Maße der zentralen Tendenz, die ein *niedrigeres* Skalenniveau voraussetzen, auch bei Daten mit einem *höheren* Skalenniveau verwenden. Man darf also beispielsweise bei intervallskalierten Daten auch den Median und den Modalwert berechnen. (Dies liegt daran, dass höherwertigere Skalen die Eigenschaften der niederwertigeren Skalen erben.)

Fassen wir zusammen:

Intervallskala: Mittelwert, Median, Modalwert

Ordinalskala: Median, Modalwert

Nominalskala: Modalwert

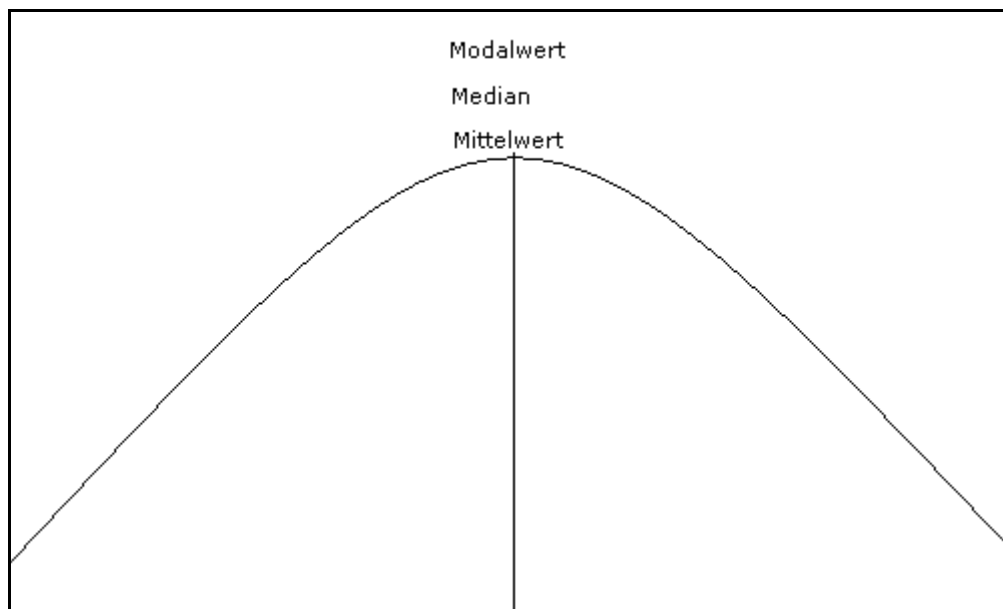
2) Vergleichen wir nun abschließend drei Formen von Verteilungen:

a) gleichverteilte Daten (symmetrische Verteilung)

b) linkssteil verteilte Daten

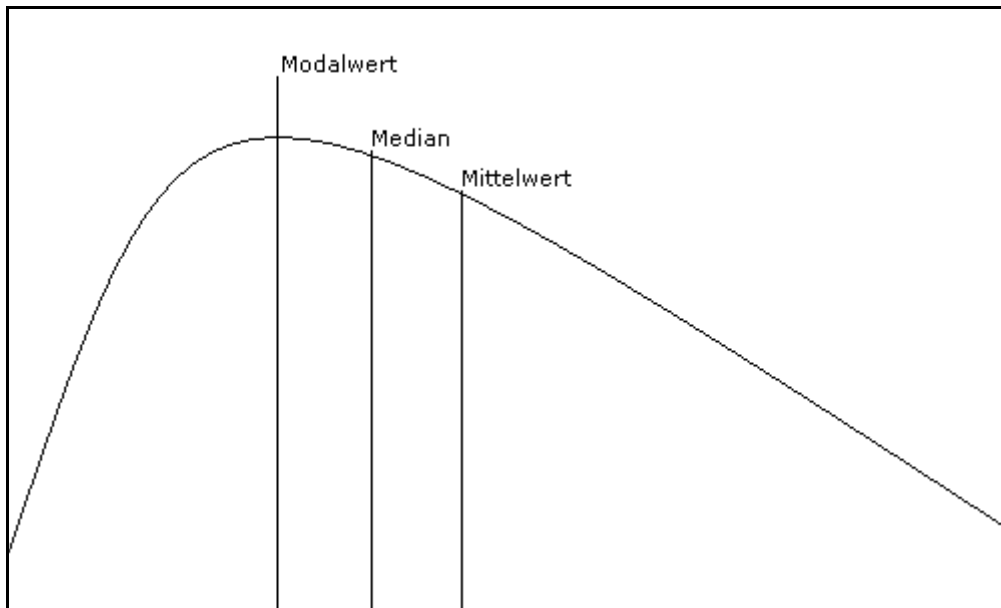
c) rechtssteile verteilte Daten

bei a) fallen Modalwert, Median und Mittelwert zusammen

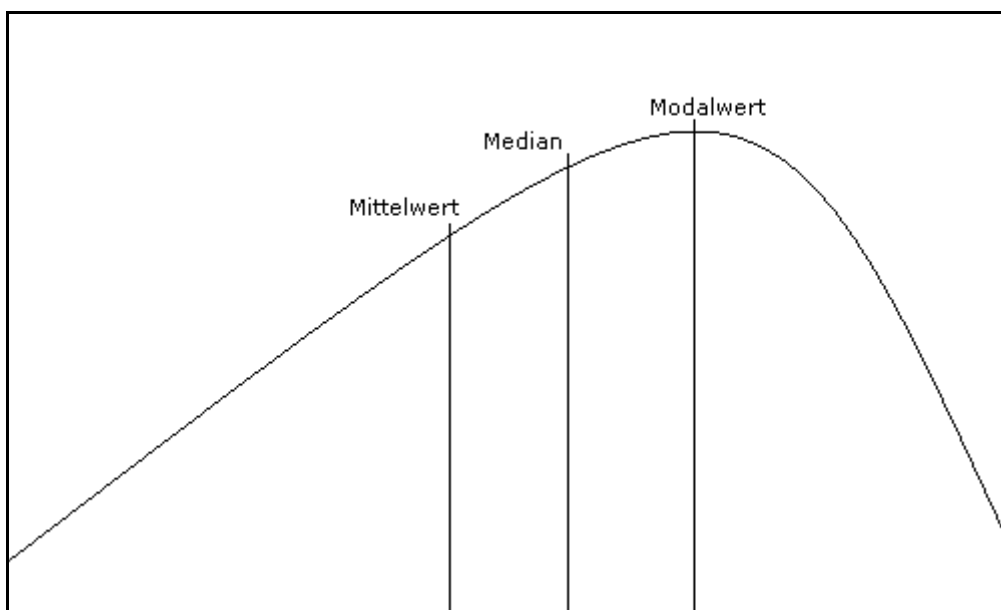


bei b) ist links der Modalwert, gefolgt vom Median, an den sich der Mittelwert anschließt





bei c) ist links der Mittelwert, gefolgt vom Median, an den sich der Modalwert anschließt.



### Dispersionsmaße

Geht es bei den Maßen der zentralen Tendenz darum, eine Gesamtverteilung aufgrund eines Wertes zu schätzen (Modalwert, Median, Mittelwert), so stellt sich bei den Dispersionsmaßen die Frage, *wie gut* eine Verteilung durch ein zentrales Tendenzmaß repräsentiert wird. Das hängt von der *Streuung* der Daten ab. Statt Dispersionsmaße können wir daher auch sagen: *Streuungsmaße*. Streunungsmaße berücksichtigen nicht nur die zentrale Tendenz, sondern geben uns auch Auskunft über etwaige 'Ausreißer'.

Streuungsmaße reflektieren das Ausmaß der *Abweichung* der einzelnen Daten von den Maßen der zentralen Tendenz.

**1) das einfachste Streuungsmaß ist die *Variationsbreite*. Diese versteht sich einfach als die Differenz des größten und des kleinsten Werts.**

In dem Beispiel unserer Daten ist die Variationsbreite:

$$41 (= \text{Maximum}) - 9 (= \text{Minimum}) = 32$$

Das Problem bei der Variationsbreite ist aber: Der größte bzw. kleinste Wert könnte ein einzelner Ausreißer sein, der wenig über die Gesamtverteilung aussagt.

**2) Deshalb ist es günstiger, den so genannten *Interquartilbereich* zu verwenden. Es handelt sich dabei um jenen Bereich, innerhalb dessen die mittleren 50% aller Fälle liegen.**

Um den Interquartilbereich zu bestimmen, benötigen wir *zwei* Werte: a) jenen Wert, der die unteren 25 % aller Fälle von den oberen 75 % abschneidet. Man bezeichnet diesen Wert auch als *erstes Quartil*.

b) jenen Wert, der die oberen 25 % aller Fälle von den unteren 75 % abschneidet. Dieser Wert wird auch als *drittes Quartil* bezeichnet.

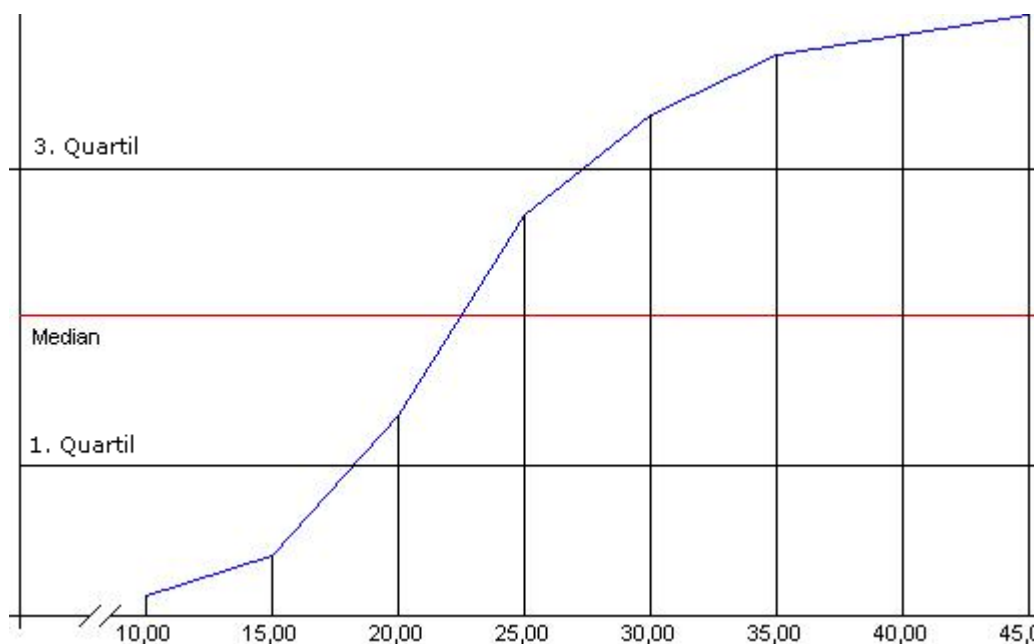
Das *zweite Quartil* ist mit dem Median identisch. Zur Erinnerung: der Median ist jener Wert, der die der Größe nach geordneten Werte in genau zwei Hälften teilt.

Zur Berechnung der Quartilwerte gehen wir ganz analog zu der Bestimmung des Medians vor:

Bei wenigen, ungruppierten Daten bestimmen wir 1. und 3. Quartil einfach durch Auszählung der der Größe nach geordneten Werte.

Bei *gruppierten* Daten gehen wir ebenfalls analog zur Medianbestimmung vor.

Man zieht in einem ersten Schritt bei 25 % eine Parallele zur x-Achse. An der Stelle, an der diese Parallele das Summenpolygon schneidet, fällt man ein senkrecht Lot zur x-Achse. Die Stelle, an der dieses senkrechte Lot die x-Achse schneidet, ist das gesuchte erste Quartil. Auf ähnliche Weise bestimmt man das 3. Quartil, nur wird hier die Parallele bei 75% zur x-Achse gezogen. Das erste Quartil beträgt in unserem Beispiel 18.21 und das dritte Quartil 27.5.



### 3a) Varianz

Denken wir zurück, wie wir weiter oben Streuungsmaße eingeführt haben: Streuungsmaße reflektieren das Ausmaß der *Abweichung* der einzelnen Daten von den Maßen der zentralen Tendenz.

Nun kommen Variationsbreite und Interquartilbereich nur unzureichend dieser Aufgabe nach. Sie sagen uns zwar etwas aus über die *generelle* Streuung der Daten, nicht aber über die Abweichung jedes *einzelnen* Wertes von dem Maß der zentralen Tendenz. Diese Aufgabe erfüllt erst die so genannte *Varianz*. Sie gilt daher innerhalb der Streuungsmaße als das *informationsreichste* Maß.

Zur Erklärung der Varianz beginnen wir mit folgender Überlegung: Wie lässt sich das Ausmaß der Abweichung vom Mittelwert bei einem einzelnen Wert rechnerisch ermitteln? Ganz einfach: Wir bilden die Differenz dieses Wertes zum Mittelwert, also:  $x_i - \bar{x}$ . Wir erhalten so die Abweichung einer einzelnen Beobachtung vom Mittelwert.

Wie können wir nun die verschiedenen Differenzen aller Werte zum Mittelwert in *eine* rechnerische Größe zusammenfassen? Man könnte versuchen, die durchschnittliche Abweichung vom Mittelwert zu berechnen, indem man die Summe aller Abweichungen durch die Gesamtanzahl dividiert. Nun ist aber, wie wir bereits wissen, die Summe aller Abweichungen vom Mittelwert Null!

Man kann nun diese Schwierigkeit umgehen, indem man jede Abweichung zuerst *quadriert*.

**Die Summe aller quadrierten Abweichungen vom Mittelwert dividiert durch die Gesamtanzahl aller Beobachtungen bezeichnet man als Varianz.**

Auf unser Beispiel bezogen bedeutet dies das folgende:

Meßwert	Differenz zum Mittelwert	quadrierte Differenz
15	-7.4333	55,2539
20	-2.4333	5,9209
17	-5.4333	29,5207
35	-12.566	157,9043
19	-3.4333	11,7875
22	-0.4333	0,1877
24	+1.566	2,4523
33	+10.566	111,6403
26	+3.566	12,7163
15	-7.4333	55,2539
41	+18.566	344,6963
22	-0.4333	0,1877
11	-11.4333	130,7203
21	-1.4333	2,0543
15	-7.4333	55,2539
25	+2.566	6,5843
24	+1.566	2,4523
21	-1.4333	2,0543
31	+8.566	73,3763
27	+4.566	20,8483
31	+8.566	73,3763
9	-13.4333	180,4535
23	+0.566	0,3203
21	-1.4333	2,0543
25	+2.566	6,5843
28	+5.566	30,9803
17	-5.4333	29,5207
13	-9.4333	88,9871
24	+1.566	2,4523
18	-4.4333	19,654

Zählt man die letzte Spalte zusammen (= 1515,249) und dividiert die Summe durch 30, so erhält man die Varianz. Diese ist in diesem Beispiel: 50.5083

in mathematischer Schreibweise :

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

### *Veränderungen der Varianz bei Lineartransformationen*

Lineartransformationen haben die allgemeine Form:  $y = b \cdot x + a$

2 Beispiele:

a)  $y = 2 \cdot x + 0$  ( $b=2$  und  $a=0$ )

Multiplizieren wir alle Messwerte mit dem Faktor 2, so ändert sich die Varianz um den Faktor  $2^2=4$  ( $=b^2$ )

Die Varianz der mit dem Faktor 2 multiplizierten Messwerte beträgt: 202.0332

Das ist  $4 \cdot 50.5083$

b)  $y = 1 \cdot x + 2$  ( $b = 1$  und  $a = 2$ )

Addieren wir zu jedem Wert der Messwertreihe die Zahl 2, so ändert sich *nicht* die Varianz.

Die Varianz der durch die additive Konstante  $+ 2$  veränderten Messwerte ist gleich der Varianz der ursprünglichen Werte, nämlich: 50.5083

Fassen wir zusammen:

Die Veränderung einer Messwertreihe durch eine additive Konstante  $a$  hat *keinen* Einfluss auf die Varianz der Messwerte. Werden die Messwerte dagegen mit einem Faktor  $b$  multipliziert, so hat die neue Messwertreihe eine um den Faktor  $b^2$  veränderte Varianz.

*Voraussetzung für die Berechnung der Varianz:*

### **Intervallskala**

### 3b) Standardabweichung

Rekapitulieren wir: die Varianz ist der Durchschnitt aller quadrierten Abweichungen vom Mittelwert. Dies entspricht aber nicht unserem ursprünglichen Maßstab, sondern einer *Quadrierung* der Einzelabstände. Um die ursprüngliche Einheit wieder zurück zu gewinnen, wird einfach die *Wurzel* aus der Varianz berechnet. Diese bezeichnet man als *Standardabweichung*.

In mathematischer Schreibweise ausgedrückt:

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

In unserem Beispiel ist die Standardabweichung: Wurzel aus 50.5083 und das ergibt ca. 7.1069

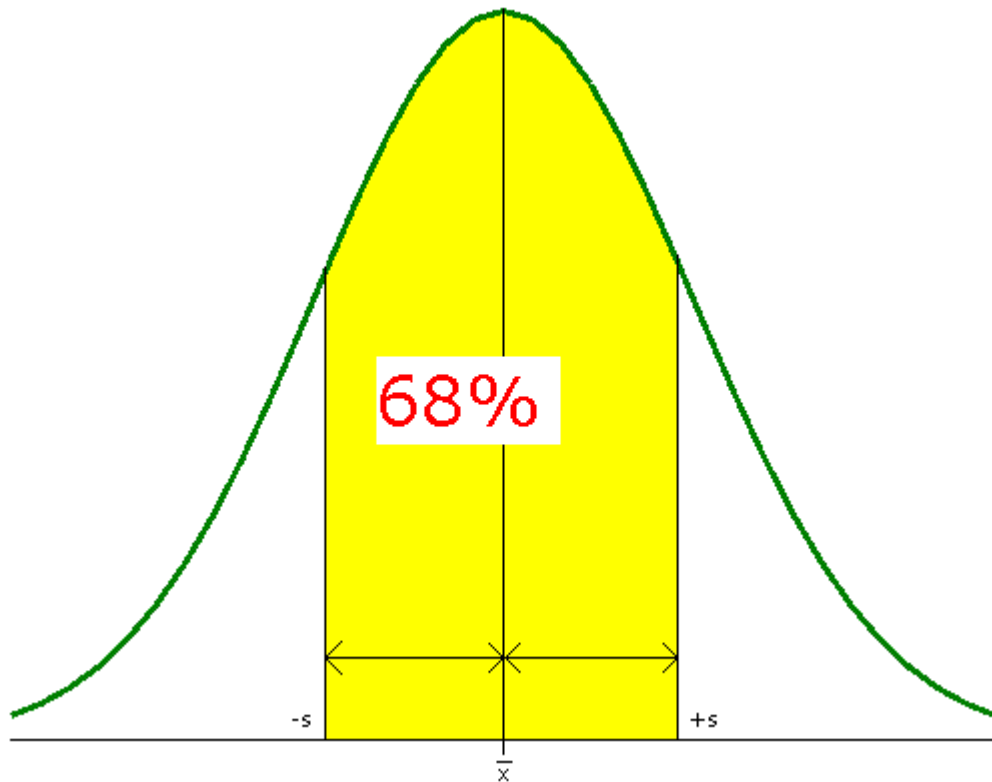
#### *Veränderungen der Standardabweichung bei Lineartransformationen*

**Multiplizieren** wir alle Messwerte mit dem Faktor b, so ändert sich die Varianz um den Faktor  $b^2$ . Da die Standardabweichung die Wurzel aus der Varianz ist, so ändert sich die Standardabweichung um den Faktor b.

**Addieren** wir zu jedem Wert der Messwertreihe die Zahl a, so ändert sich die Varianz nicht und somit auch nicht die Standardabweichung..

#### *Eigenschaften der Standardabweichung bei Normalverteilung*

Nehmen wir an die Verteilung unserer Daten sei *unimodal* (eingipflig) und *symmetrisch* und habe dazu noch einen *glockenförmigen* Verlauf. In diesem Falle spricht man von einer *Normalverteilung*.



Eine Normalverteilung hat nun die Eigenschaft, dass in dem Bereich  $\bar{x} - 1 \cdot s$  und  $\bar{x} + 1 \cdot s$  ca. 68 % aller Werte anzutreffen sind. D.h.:

*Ziehen wir aus unseren Messwerten rein zufällig einen Wert, so werden wir in ca. 68% aller Ziehungen bzw. mit einer Wahrscheinlichkeit von ca. 0,68 einen Wert zwischen  $\bar{x} - 1 \cdot s$  und  $\bar{x} + 1 \cdot s$  bekommen.*

Außerhalb von  $+1s$  bzw.  $-1s$  liegen jeweils 16 % aller Messwerte.

Zwischen  $\bar{x} - 2 \cdot s$  und  $\bar{x} + 2 \cdot s$  liegen ca. 95% aller Fälle.

In unserem Beispiel liegen zwischen  $22.43 + 1 \cdot 7.1069$  und  $22.43 - 1 \cdot 7.1069$  68 % aller Werte - vorausgesetzt unsere Verteilung ist eine Normalverteilung!

#### z-Werte:

Mittelwert und Standardabweichung dienen dazu, um Werte in verschiedenen Kollektiven miteinander vergleichen zu können.

Stellen Sie sich vor, eine Person hat in einem Kollektiv von 30 Personen einen Wert von 60 bekommen (beispielsweise die Anzahl von Punkten).

Dieser Wert 60 sagt uns wenig, was er wirklich bedeutet. Ist er hoch, niedrig, mittel? Das hängt von der Verteilung der Werte in der ganzen Gruppe ab.

So besteht ein Unterschied, ob der Wert 60 aus einem Kollektiv mit dem Mittelwert 30 und einer Standardabweichung von 10 gezogen wurde oder aus einem Kollektiv mit einem Mittelwert von 65 und einer Standardabweichung von 20.

Um die relative Position des Wertes in dem jeweiligen Kollektiv feststellen zu können, berechnet man für jeden Wert einen sogenannten standardisierten z-Wert nach der Formel:

$$z = \frac{x_i - \bar{x}}{s}$$

Dieser z-Wert sagt uns, wie viele Standardabweichungen der jeweilige Wert von dem Mittelwert im Kollektiv abweicht.

Charakteristisch für die z-transformierten Werte ist: Ihr Mittelwert ist 0 und ihre Standardabweichung ist 1

Bezogen auf das hier verwendete Datenbeispiel bedeutet dies:

V1	Z
15	-1.05
20	-.34
17	-.77
35	1.77
19	-.48
22	-.06
24	.22
33	1.49
26	.50
15	-1.05
41	2.62
22	-.06
11	-1.61
21	-.20
15	-1.05
25	.36
24	.22
21	-.20
31	1.21
27	.64
31	1.21
9	-1.89
23	.08
21	-.20
25	.36
28	.78
17	-.77
13	-1.33
24	.22
18	-.62

Die Summe der z-Werte ergibt (bis auf Rundungsfehler) 0 und eine Standardabweichung von 1.



Frage: Warum ist das so?