4 Statistische Maßzahlen
(Kennwerte und Parameter)

4.1 Kennwerte der Lokalisation und der Variation

Thema dieses Kapitels ist die numerische Charakterisierung von Daten. Darunter verstehen wir eine zusammengefasste Darstellung einer Datengruppe durch Kennwerte der Lokalisation, der Variation (Dispersion), im weiteren Sinne auch der Korrelation und der Regression. Das folgende Beispiel zeigt dies für Kennwerte der Lokalisation und Variation.

Beispiel 1
Bei geschälten Sonnenblumenkernen wurde der Massenanteil w(Fett) per Soxhletextraktion bestimmt. Die errechneten Massenanteile wurden auf eine Nachkommastelle (in g) genau gerundet (Inkrement = 0,1 g/100 g) und notiert. Bei 25 Proben erhielten wir so folgenden Werte für w(Fett) in g/100 g.(Unter Inkrement verstehen wir den kleinsten Betrag um den ein Zahlenwert schrittweise verändert werden kann.

50.0	49.5	50.4	50.4	50.4	50.4	49.8	49.9	50.1	50.8
50.4	50.3	50.0	50.3	50.5	50.3	51.1	50.3	50.3	49.8
49.9	50.6	50.1	49.9	50.1					

Tabelle 1

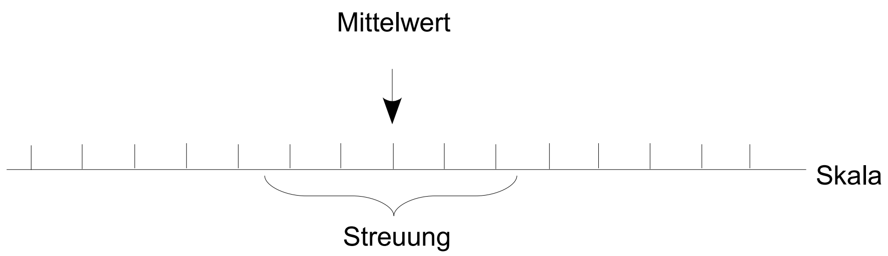
Wenn wir zur Verbesserung der Übersicht die 25 Werte durch <u>einen</u> repräsentativen Wert darstellen wollen, dann lässt sich das durch Betrachten der Tabelle nicht unmittelbar erreichen. Zu diesem Zweck berechnen wir den Mittelwert und haben damit ein Charakteristikum zur Beschreibung der 25 Ergebnisse durch <u>einen</u> Wert, nämlich w(Fett) = 50,22 g/100 g. Wenn wir nach einem Wert fragen, der etwas über die Schwankungen der Einzelwerte um den Mittelwert aussagt, dann berechnen wir dazu z.B. die Standardabweichung (s_x) der Einzelwerte vom Mittelwert. Für diese gilt hier s_x = ± 0,34 g/100 g und das bedeutet, im Bereich Mittelwert ± Standardabweichung liegen ca. 68 % aller Werte (darauf gehen wir später genauer ein). Ist dieser Bereich eng, dann streuen die Werte schwach, ist er weit, dann streuen sie stark.

Die Präsentation der Liste mit 25 Einzelwerte kann nun durch Angabe von Mittelwert und Streuung mit nur zwei Zahlen aussagekräftig und übersichtlich unterstützt werden.

$$\begin{aligned} &\text{Mittelwert } \bar{x} &&= \quad 50{,}22 \text{ g/100 g} \\ &\text{Standardabweichung } s_x &&= \pm 0{,}34 \text{ g/100 g} \end{aligned} \quad \left(\begin{array}{l} \text{gesprochen x quer} \\ \text{gesprochen s x} \end{array} \right)$$

Wir nennen diese beiden Werte Kennwerte oder Maßzahlen einer Datenreihe. Da es sich dabei nicht immer um Zahlen handelt, werden wir in der Folge nur von Kennwerten sprechen.

Der Mittelwert kennzeichnet die punktuelle Lage des charakteristischen Wertes auf einer Skala, der Streuungswert dagegen auf der Skala <u>einen Bereich</u>. (Abb.1)



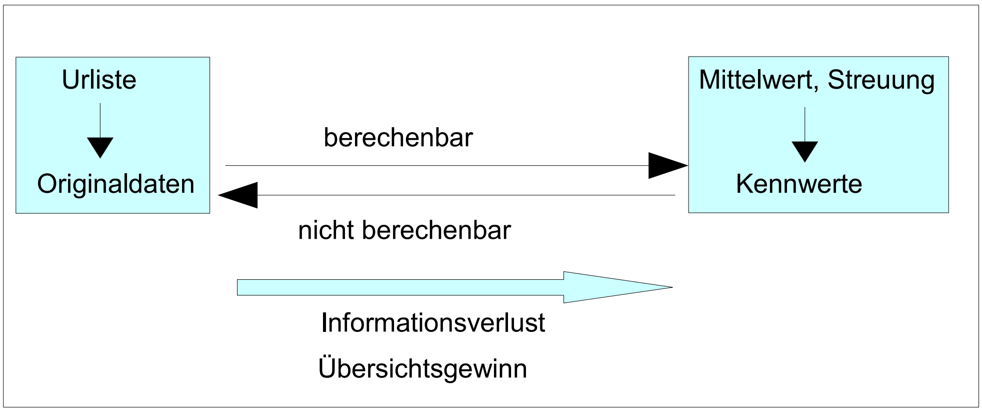
Mittelwerte → Kennwerte der Lokalisation
Streuungswerte → Kennwerte der Variation.

Informationsverlust

Durch die Zusammenfassung (Aggregation) der 25 Werte zu zwei Kennwerten gewinnen wir an Übersicht. Die beiden Kennwert sagen auf einen Blick in Kürze mehr als die Tabelle. Diesem Gewinn an Übersicht steht aber immer ein Informationsverlust gegenüber. So informiert der Mittelwert über die "zentrale Tendenz" der 25 Werte, über die Größe der Extremwerte und der Einzelwerte sagt er aber nichts. Die Standardabweichung gibt zwar Auskunft über die mittlere Abweichung der Einzelwerte vom Mittelwert, über die wirklichen Abstände der Einzelwerte voneinander sagt er aber auch nichts.

Während wir aus den Originaldaten immer die Kennwerte ermitteln können, ist der umgekehrte Weg nicht mehr möglich. So können wir aus den fünf Werten 15; 20; 21; 26 und 33 den Mittelwert 23 berechnen. Aus diesem Mittelwert 23 und der Anzahl der Werte = 5 können wir zwar die Summe der Einzelwerte und dann beliebige fünf Werte berechnen, die den Mittelwert 23 haben, z.B. 1; 27; 44; 12 und 31, nicht aber gezielt die fünft originären Werte. Wegen dieses Verlustes an Information gilt:

Urliste immer aufbewahren.

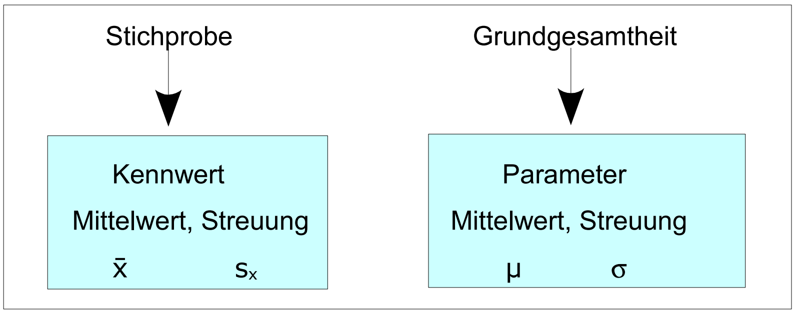


Kennwerte - Parameter

Kennwerte werden meist aus Daten von Stichproben ermittelt und sind damit Schätzwerte für die Daten der Grundgesamtheit. Die den Kennwerten entsprechenden Werte der Grundgesamtheit nennen wir Parameter. Der Stichprobenmittelwert $\bar{x}$ ist als Kennwert eine Schätzung für den Parameter μ, den Mittelwert der Grundgesamtheit.

Folgende Vereinbarung ist üblich:

Parameter → kleine **griechische** Buchstaben
Kennwerte → kleine **lateinische** Buchstaben



μ gesprochen mü (das kleine griechische m)
σ gesprochen sigma (das kleine griechische s)

4.2 Kennwerte der Lokalisation

Die meisten von uns werden unter dem Begriff Mittelwert das verstehen was weiter unten als arithmetisches Mittel beschrieben wird. Neben diesem gibt es aber auch noch andere Mittelwerte. Wir werden uns hier mit den folgenden beschäftigen

Arithmetisches Mittel
Geometrisches Mittel
Harmonisches Mittel
Modalwert
Medianwert
Quantile

Welcher dieser Mittelwerte im konkreten Fall angewendet wird, das hängt von der Art und Verteilung der Daten sowie von der Fragestellung ab.

Bei den folgenden Berechnungen werden Messwerte in Formeln eingesetzt. Während wir beim Rechnen mit physikalischen Größengleichungen zu den Zahlen immer die Einheiten angeben, ist das bei statistischen Berechnungen nicht üblich. Wir setzen nur die Zahlenwerte in die Gleichungen ein.

4.2.1 Das arithmetische Mittel ($\bar{x}$)

Wenn wir vom Mittelwert sprechen, dann meinen wir in der Regel das arithmetische Mittel. Sind Verwechslungen mit anderen Mittelwerten denkbar, dann sollten wir immer angeben, dass es sich konkret um das arithmetische Mittel handelt.

Eigenschaften:	$\Sigma(\bar{x}-x_i) = 0$ $\Sigma x_i = n * \bar{x}$ Das arithmetische Mittel muss in der Datenreihe aus der es berechnet wurde, nicht realisiert sein (d.h. nicht selber vorkommen). Es wird stark von Extremwerten beeinflusst.

Anwenden:	bei quantitativen, metrisch skalierten Daten
	    bei Intervalldaten
	    bei Proportionaldaten = Verhältnisdaten
	streng genommen nur bei symmetrischen Verteilungen

Nicht anwenden:	bei Nominaldaten bei Ordinaldaten bei multimodalen Verteilungen bei stark asymmetrischen Verteilungen bei offenen Randklassen bei extrem kleinen Stichproben

Das folgende Beispiel zeigt den an sich banalen Formalismus der Berechnung.

Beispiel 2
Bei einem Diabetiker wurde im Laufe mehrerer Jahren alle 3 Monate der HbA1c-Wert gemessen, der den %-Satz des glykierten Hämoglobins im Blut angibt. Wir begnügen uns hier mit 10 Werten aus einer umfangreichen Urliste.

i	1	2	3	4	5	6	7	8	9	10
$\mathbf{x_i}$	**8.8**	**7.0**	**7.0**	**5.7**	**6.3**	**6.9**	**7.3**	**6.6**	**6.9**	**6.5**

Tabelle 2

Das arithmetische Mittel wird berechnet, indem wir die Summe der Einzelwerte (Σx_i) durch die Anzahl (n) der Einzelwerte dividieren.

Notationen (hier die Schreibweise mathematischer Formulierungen) zur Statistik sind in der Literatur bedauerlicherweise uneinheitlich. Wir werden die hier verwendeten Notationen bei Einführungen jeweils vorstellen.

Es gelten

$\bar{x}$	= arithmetisches Mittel	z.B. 6,90 %
x_i	= Einzelwert bei i-ter Messung	z.B. 5,7 %
i	= Laufindex	z.B. 4
x_4	= 4. Einzelwert	5,7 %
n	= Anzahl der Einzelwerte	10
Σ	= großes griechisches S, gelesen Sigma, bedeutet: Summiere die hinter dem Zeichen stehenden Werte.	
$\sum_{i=1}^{n} x_i$	= bedeutet: Summieren Sie alle Werte von i = 1 bis i = n. gesprochen: Summe über x_i von i = 1 bis i = n. Wenn bei der Notation Verwechslungen ausgeschlossen sind, dann schreiben wir nur Σx_i.	

Berechnet wird $\bar{x}$ nach

Arithmetisches Mittel

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{69}{10}$$

$$\bar{x} = \frac{1}{n} \Sigma x_i = \frac{1}{10} * 69$$

$$\bar{x} = \Sigma \frac{xi}{n} = \Sigma\, 0{,}88+0{,}70+0{,}70+0{,}57+0{,}63+0{,}69+0{,}73+0{,}66+0{,}69+0{,}65 = \mathbf{6{,}9}$$

In der Regel wir das arithmetische Mittel mit einer Dezimalstelle mehr als die vorliegenden Daten angegeben.

Beispiel 3
Nehmen wir an, es lägen von der Urliste für Beispiel 2 die folgenden 10 Messwerte vor, die aber mit unterschiedlicher Stellenzahl nach dem Komma angegeben wären, etwa so

Messwert	8.8	**7.04**	7.0	**5.69**	**6.29**	6.9	7.3	6.6	6.9	6.

Tabelle 3

Der Grund dafür könnte sein, dass die Werte 2; 4 und 5 von einem anderen Labor ermittelt wurden, welches die Zahlen mit zwei Nachkommastellen lieferte. Das Labor, welches die restlichen Daten bereitstellte, hatte die Daten vor der Datenübergabe auf eine Nachkommastelle gerundet. Die ungerundeten Werte dieser Daten kennen wir also nicht. Wie setzen wir diese unterschiedlich genauen Daten in die Berechnung des Mittelwertes ein? Die Werte 2; 4 und 5 liegen mit einer höheren Genauigkeit (Inkrement 0,01 %) vor als die Übrigen (Inkrement 0,1 %). In einer solchen Situation gehen gehen wie folgt vor. Wir runden die drei genaueren Werte nach DIN 1333 (bei 0 bis <5 → abrunden; bei 5 bis 9 → aufrunden), so dass sie der Stellenzahl des ungenauesten Wertes der Daten entsprechen. Daraus folgt 7,04 → 7,0; 5,69 → 5,7; 6,29 → 6,3. Mit den gerundeten Werten wird dann wie üblich das arithmetische Mittel berechnet.

Hinweis: Manchmal können wir Berechnung vereinfachen, wenn wir folgendes Berücksichtigen. Wenn zu jedem x_i-Wert der Daten die gleiche Zahl addiert (subtrahiert) wird, so wächst (vermindert sich) $\bar{x}$ um den gleichen Wert. Das Entsprechende gilt für die Multiplikation (Division) der x_i-Werte. Hier ändert sich $\bar{x}$ um den gleichen Faktor. Stellen wir uns vor, wir müssten zur Addition die folgenden Werte in einen Taschenrechner eintippen

0,00353
0,00313
0,00267
0,00363

um den Mittelwert zu bilden. $\Sigma x_i = 0{,}01296$; $\bar{x} = 0{,}00324$.

Wenn wir vor dem Eintippen jeden Wert mit 10^5 multiplizieren, dann sparen wir das Eintippen der Nullen und des Kommas.

353
313
267
363

Dann ist $\Sigma x_i = 1296$ und $\bar{x} = 324$. Wenn wir nun 324 mit dem Faktor 10^{-5} multiplizieren, dann erhalten wir 0,00324. Bei längeren Datenreihen können wir durch solche Verfahren Zeit sparen.

Das arithmetische Mittel, wenn einzelne Messwerte gehäuft vorkommen

Beispiel 4
Einer retrospektiven Erhebung entstammen folgende 120 Werte der systolischen Blutdrucks einer 34 jährigen Frau. Die Daten sind Teil einer umfangreicheren Urliste mit Messwerten der gleichen Person zwischen 96 und 175 mm Hg. (Häufigkeit = H_i)

H_i	9	9	6	6	21	10	16	25	18
mm Hg	114	115	116	117	118	119	120	121	122

Tabelle 4

Es fällt auf, das wir hier neun verschiedene Messwerte (114 bis 122) haben, die unterschiedlich häufig vorkommen. Wenn wir diese neun Messwerte addieren und durch neun dividieren, erhalten wir 1062/9 = 118,0. Dieser Wert liegt zwar im mittleren Bereich der geordneten Messwertreihe, es ist aber leicht zu erkennen, dass er die Messwerte nicht gut repräsentiert. Denn im rechten Teil der Reihe liegen, wie wir den Häufigkeiten entnehmen, deutlich mehr Werte als im linken Teil. Der Mittelwert müsste also höher liegen. Wie berechnen wir in einer solchen Situation den Mittelwert? Wir müssen bei der Berechnung die Häufigkeiten der einzelnen Messwerte berücksichtigen. Es sind zusammen $\sum H = 120$ Einzelwerte. Wir könnten alle 120 Werte der Urliste addieren und die Summe durch 120 dividieren. Die Urliste mit den 120 Werten liegt uns aber nicht vor. Wenn wie hier einzelne Daten gehäuft vorkommen, und sie mit ihren Häufigkeiten tabelliert sind, dann bietet sich eine günstigere Berechungsform an. Dazu multiplizieren wir jeden Messwert x_i mit seiner Häufigkeit H_i wie das in der vorliegenden Liste schon geschehen ist.

H_i	9	9	6	6	21	10	16	25	18
x_i mm Hg	114	115	116	117	118	119	120	121	122
$x_i * H_i$	1026	1035	696	702	2478	1190	1920	3025	2196

Tabelle 5

Nun berechnen wir das arithmetische Mittel nach

Arithmetisches Mittel bei häufig vorkommenden Werten

$$\bar{x} = \frac{\sum_{i=1}^{k} (x_i * H_i)}{\sum_{i=1}^{k} H_i}$$

Notation: k = Anzahl der verschiedenen Messwerte, hier 9

$$\bar{x} = \frac{\sum_{i=1}^{k} (H_i * x_i)}{\sum_{i=1}^{k} fH_i} = \frac{114*9 + 115*9 + 116*6 + 117*6 + 118*21 + 119*10 + 120*16 + 121*25 + 122*18}{9 + 9 + 6 + 6 + 21 + 10 + 16 + 25 + 18}$$

$$\bar{x} = \frac{14268}{120} = 118{,}9$$

Diese Rechnung sieht auf den ersten Blick komplizierter aus, sie ist aber letztlich mit weniger Aufwand verbunden. Natürlich geht mit Excel alles schneller. Wir sehen, dass der Mittelwert $\bar{x}$ = 118,9 wie vermutet, höher liegt als $\bar{x}$ = 118,0 nach der falschen Berechnung.

Das gewogene arithmetische Mittel

Beispiel 5

Aus Zuchtbehälter 1 haben wir 50 Larven des Mehlkäfers (Tenebrio molitor) einzeln gewogen. $\bar{x}_1$ = 302 mg. Aus Zuchtbehälter 2 wurden 125 Tiere gewogen. $\bar{x}_2$ = 285 mg. Wir interessieren uns nun für das mittlere Gewicht aller 175 Tiere und bilden dazu den Mittelwert der beiden Mittelwerte nach ($\bar{x}_1$ + $\bar{x}_2$)/2 = 587/2 = 293,5 mg. Die beiden Mittelwerte, die wir hier gleichwertig behandelt haben, sind aber nicht gleichwertig, da $\bar{x}_2$ aus einer deutlich größeren Stichprobe ermittelt wurde (n = 125) als $\bar{x}_1$ (n = 50). $\bar{x}_2$ repräsentiert also die Grundgesamtheit in Behälter 2 besser als $\bar{x}_1$ die Grundgesamtheit in Behälter 1. Wir müssen, um das zu berücksichtigen, die beiden Mittelwerte ihren Stichprobenumfängen entsprechend gewichten. Dies geschieht durch folgende Berechnung.

$$\bar{x}_{gesamt} = \frac{\bar{x}_1 * n_1 + \bar{x}_2 * n_2}{n_1 + n_2} = \frac{302 * 50 + 285 * 125}{175} = 50725/175 = 289{,}9$$

Hierdurch geht der Mittelwert $\bar{x}_2$ also 125 mal in die Berechnung ein und $\bar{x}_1$ nur 50 mal.

Die allgemeine Formel lautet:

Arithmetisches Mittel gewogen

$$\bar{x} = \frac{\sum_{i=1}^{k} (\bar{x}_i * n_i)}{\sum_{i=1}^{k} n_i}$$

k = Anzahl der Stichproben (hier 2)
n_1 = Umfang der Stichprobe 1 (hier 50)
n_2 = Umfang der Stichprobe 2 (hier 125)

4.2.2 Geometrisches Mittel ($\bar{x}_G$)

Eigenschaften: $\bar{x}_G < \bar{x}$

Anwenden: wenn eine Variable, (Zellzahl einer Zellkultur) sich in Abhängigkeit von z.B. der Zeit (pro Stunde) nicht linear ändert. Das liegt vor, wenn die Daten auf semilogarithmischem* Papier angenähert eine Gerade ergeben (sich exponentiell ändern).
bei absolutskalierten Daten
bei logarithmisch normalverteilten Daten
nur bei positiven Daten

Nicht anwenden: wenn $x_i = 0$
wenn x_i negativ
bei multimodalen Verteilungen

* semilogarithmisches Papier: Abszissen metrisch, Ordinate logarithmisch geteilt.

Beispiel 6

Zur Einführung soll ein <u>fiktives</u> Beispiel aus der Mikrobiologie zeigen, dass hier das arithmetische Mittel zu einem falschen Ergebnis führt. Nehmen wir an, die Zellzahl einer Mikroorganismenkultur hätte sich in den letzten vier Tagen so entwickelt:

Anfang des	1. Tages :	1000 Zellen/mL
Ende des	1. Tages :	4000 Zellen/mL
Ende des	2. Tages :	20000 Zellen/mL
Ende des	3. Tages :	60000 Zellen/mL
Ende des	4. Tages :	240000 Zellen/mL

Nach Auftragen der Zellzahl/mL gegen die Zeit resultieren Kurven, die auf exponentielles Wachstum schließen lassen. (Abb. 1 und Abb. 2)

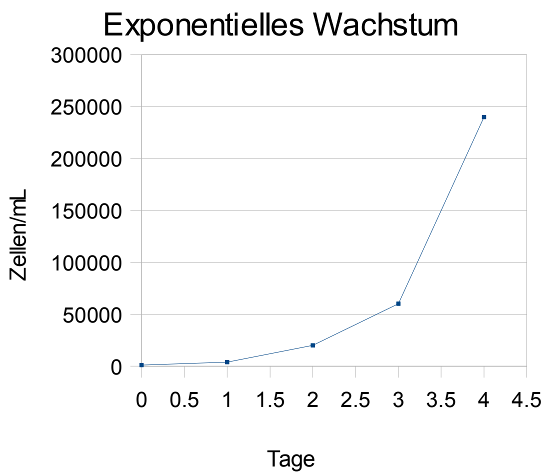


Abb.1

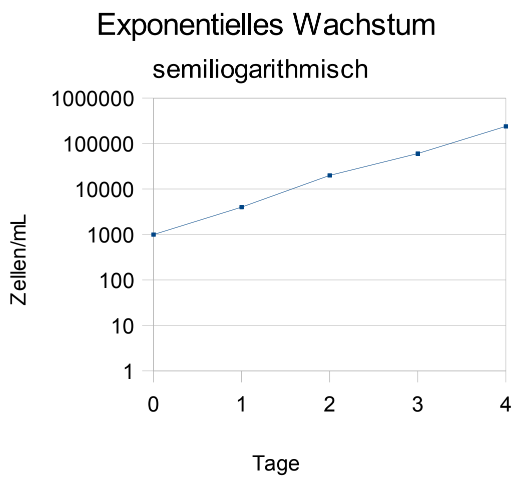


Abb.2

Die Zellzahl wächst, wie die Graphik zeigt, nicht mit der Zeit proportional.

Der Vermehrungsfaktor der MO am ersten Tag wäre 4000/1000 = 4
Der Vermehrungsfaktor der MO am zweiten Tag wäre 20000/4000 = 5
Der Vermehrungsfaktor der MO am dritten Tag wäre 60000/20000 = 3
Der Vermehrungsfaktor der MO am vierten Tag wäre 240000/60000 = 4

Die unterschiedlichen Vermehrungsfaktoren mögen durch wechselnde Umweltfaktoren für die Kultur begründet sein. Wir wollen nun fragen, wie groß ein gleichbleibender täglicher Vermehrungsfaktor sein müsste, um am Ende des 4. Tages 240000 Zellen/mL zu erreichen. Mitteln wir die vier Faktoren (4;5;3,4) arithmetisch, so erhalten wir als mittleren Vermehrungsfaktor 4. Wenn wir damit die Zellzahlen am Ende eines jeden Tages berechnen, dann erhalten wir:

1000	* 4	=	4000
4000	* 4	=	16000
16000	* 4	=	64000
64000	* 4	=	256000 (und nicht 240000)

Das arithmetische Mittel der Vermehrungsfaktoren ist also falsch, denn es führt nicht zum richtigen Ergebnis. Führen wir die gleiche Überprüfung mit dem Faktor 3,93597 durch, so kommen wir zum korrekten Ergebnis.

1 000	* 3,93597	=	3935,97
3935,97	* 3,93597	=	15491,86
13416,39687	* 3,93597	=	60975,50
60975,50	* 3,93597	=	239997,72 (Abweichung von 240000 durch Rundung bedingt)

Dieser Faktor ist also richtig. Er ist das geometrische Mittel der vier Faktoren. Wie es berechnet wird, das zeigt das folgende Beispiel aus der Mikrobiologie mit realen Werten.

Beispiel 7

Bei einer Kultur von Escherichia coli K12 wurde die Keimzahl mit dem Kochschen Plattengussverfahren über vier Stunden stündlich untersucht. Beim Start der Untersuchung, also zu Beginn der ersten Stunde, war die Keimzahl 5,8 * 10^3 K/mL . Die weiteren Zählergebnisse und die Vermehrungsfaktoren finden wir in der folgenden Tabelle. Zu berechnen ist der mittlere Vermehrungsfaktor, also der Faktor, der, wenn er über die vier Stunden gleich geblieben wäre, zu der Keimzahl 268,3*10^3 Zellen/mL geführt hätte.

Beginn	K/mL	Ende	K/mL	Vermehrungsfaktor
1. Stunde	5,8*10^3	1. Stunde	12,8*10^3	2.21
2. Stunde	12,8*10^3	2. Stunde	34,6*10^3	2.7
3. Stunde	34,6*10^3	3. Stunde	107,3*10^3	3.1
4. Stunde	107,3*10^3	4. Stunde	268,3*10^3	2.5

Tabelle 6

Der mittlere Vermehrungsfaktor ist das geometrische Mittel der vier Vermehrungsfaktoren. Zur Berechnung des geometrischen Mittels können folgende Formeln angewendet werden.

$$\bar{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Geometrisches Mittel

$$\bar{x}_G = \text{num log } \{(\sum \log x_i)/n\}$$

Die zweite Formel ist vor allem dann interessant, wenn viele große Zahlen zu multiplizieren sind. Das ist logarithmisch günstiger zu rechnen.

Notation:

$\sqrt[n]{\prod_{i=1}^{n} x_i}$ bedeutet: Multiplizieren Sie alle x_i-Werte von i = 1 bis i = n
∏ ist das große griechische P gesprochen pi
n = Anzahl der zu mittelnden Werte
i = Laufindex

$\bar{x}_G$ = geometrisches Mittel
log = dekadischer Logarithmus
num log = Numerus des dekadischen Logarithmus

Zunächst die Berechnung über die Logarithmen

$\bar{x}_G$ = num log {(∑ log x_i)/n}
$\bar{x}_G$ = num log {1.66505/4}
$\bar{x}_G$ = num log 0.41626
$\bar{x}_G$ = 2,608 → 2,61

log 2,21 = 0.34439
log 2,70 = 0.43136
log 3,10 = 0.49136
log 2,50 = 0.39794

∑ log x_i = 1.66505

Das geometrische Mittel der Vermehrungsfaktoren beträgt 2,61. Die Überprüfung bestätigt das:

$5{,}8 \cdot 10^3$	* 2,61	= $15{,}1 \cdot 10^3$
$15{,}1 \cdot 10^3$	* 2,61	= $39{,}5 \cdot 10^3$
$39{,}5 \cdot 10^3$	* 2,61	= $103{,}62 \cdot 10^3$
$103{,}62 \cdot 10^3$	* 2,61	= $269{,}15 \cdot 10^3$ (Abweichung von $268{,}3 \cdot 10^3$ durch Rundung)

Und jetzt die Berechnung durch Radizieren.

$$\bar{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

$$\bar{x}_G = \sqrt[4]{\prod_{i=1}^{4} x_i}$$

2,21 * 2,70 * 3,10 * 2,50 = 46,24425

$$\bar{x}_G = \sqrt[4]{46{,}244}$$

$$\underline{\bar{x}_G = 6{,}21}$$

Ohne Taschenrechner oder ein Rechenprogramm müssten wir die 4. Wurzel logarithmisch berechnen, was letztlich auf die vorherige Rechnung hinausläuft.

$\sqrt[4]{46{,}244}$	= (log 46,244)/4
log 46,244	= 1.66505
1.66505/4	= 0.41626
num log 0.41626	= 2,61

4.2.3 Harmonisches Mittel ($\bar{x}_H$)

Anwenden	bei Quotienten mit gleichem Nenner (20 km/h, 45 km/h) bei Intervaldaten bei Absolutdaten Überlebenszeiten mit ∞ großen Werten (1/∞ = 0)
Nicht anwenden	wenn x_i = 0

Beispiel 8
Im Zusammenhang mit mikroskopischen Untersuchungen zur Zellstruktur höherer Pflanzen haben wir in fünf Zellen der Staubfädenhaare von Tradescantien die Geschwindigkeit der Protoplasmaströmung gemessen. Dies geschah mit Hilfe eines Objektmikrometers und einer Stoppuhr. An jeder Zelle wurde

eine Messung durchgeführt. Für die einführende Erklärung des harmonischen Mittels verwenden wir nur die Ergebnisse der Messungen 1 und 2. Bei Messung 1 haben wir festgestellt, dass das Plasma die Strecke von 100 µm in 47,6 s zurückgelegt hat und die 2. Messung ergab für die gleiche Strecke 22,2 s. Daraus resultieren folgende Geschwindigkeiten:

Messung 1: In 47,6 s → 100 µm
In 1 s → 2,1 µm → v_1 = 2,1 µm/s

Messung 2: In 22,2s → 100 µm
In 1 s → 4,5 µm → v_2 = 4,5 µm/s

Geschwindigkeiten sind Quotienten mit gleichem Nenner (hier 1 s) die sich durch die Zähler (hier 2,1 µm und 4,5 µm) unterscheiden.

Wenn wir nach der mittleren Geschwindigkeit v_m fragen und das arithmetische Mittel nach $\bar{x} = (1/n)\ \Sigma\ xi = 0{,}5 * (2{,}1 + 4{,}5) = 0{,}5 * 6{,}6 = 3{,}3$ µm/s berechnen, so ist, wie die folgende Überprüfung zeigt, dieser Wert falsch.

Bei den beiden Bewegungen wurden 100 µm + 100 µm = 200 µm in 47,6 s + 22,2 s = 69,8 s zurückgelegt.

Also gilt: in 69,8 s → 200 µm
in 1 s → 2,87 µm → v = 2,87 µm/s. Das arithmetische Mittel 3,3 µm/s stimmt also nicht.

Wenn Quotienten mit gleichem Nenner zu mitteln sind, muss als Mittelwert das harmonische Mittel ($\bar{x}_H$) nach der folgenden Formel berechnet werden

Harmonisches Mittel $$\bar{x}_H = \frac{n}{\sum_{i=1}^{n} (1/x_i)}$$

$$\bar{x}_H = \frac{2}{1/2{,}1 + 1/4{,}5} = \frac{2}{(4{,}5 + 2{,}1)/9{,}45} = \frac{2}{6{,}6/9{,}45} = 2{,}864$$ (Hauptnenner suchen!)

$\bar{x}_H$ = 2,864 µm/s

Dieser Wert stimmt (abgesehen von der Rundung) mit der obigen Prüfung 2,87 µm überein.

Wir wollen nun die mittlere Geschwindigkeit für die fünf Messergebnisse berechnen.

v_1 = 2,1 µm/s
v_2 = 4,5 µm/s
v_3 = 2,5 µm/s
v_4 = 3,4 µm/s
v_5 = 2,9 µm/s

$$\bar{x}_H = \frac{n}{\sum_{i=1}^{n} (1/x_i)} = \frac{5}{1/2{,}1 + 1/4{,}5 + 1/2{,}5 + 1/3{,}4 + 1/2{,}9} = 2{,}89$$

<u>Die mittlere Geschwindigkeit beträgt also 2,89 µm/s</u>

Anwendungen des harmonischen Mittels sind neben der Berechnung mittlerer Geschwindigkeiten (Weg/Zeiteinheit) die Berechnung mittlerer Stückkosten bei Waren (€/Stück) und in der Biologie die Berechnung von mittleren Überlebenszeiten. Siehe hierzu Übung Nr.4

4.2.4 Der Modalwert (D) (Modus, Dichtemittel)

Eigenschaften	Stabil gegen Extremwerte bei Nominaldaten einziger Lagekennwert Ohne Berechnung leicht zu ermitteln
Anwenden	bei quantitativen Daten Intervalldaten Verhältnisdaten bei Nominaldaten bei Ordinaldaten bei multimodalen Verteilungen nur bei genügend großen n
Nicht anwenden	wenn Extremwerte wichtig sind, etwa bei Messungen im Zusammenhang mit der Druckfestigkeit von Behältern. Hier dürfen Extrema nicht unbeachtet bleiben.

Der Modalwert ist der Wert einer Gruppe von Daten, der am häufigsten auftritt. Wenn in einer Gruppe mehrere Maxima, deren Häufigkeiten nicht gleich sein müssen, vorkommen, dann haben wir mehrere Modalwerte (Siehe bei Häufigkeitsverteilungen). Der Modus wird nicht berechnet, sondern durch Vergleich der Häufigkeiten ermittelt.

Modalwert D = häufigster Wert des Datenbereichs

Beispiel 9
Modalwert bei quantitativen Daten
Aus einer Zucht von Schwarzkäfern (Zophobas morio) haben wir Larven auf 10 mg genau gewogen, die Messwerte der Größe nach geordnet und dann deren Häufigkeit bestimmt.

Messwert x_i in mg	540	550	560	570	580	590	600
Häufigkeit H_i	3	6	6	12	1	6	4

Tabelle 7

Ohne jede Rechnung ist aus der geordneten Liste sofort zu erkennen, dass der Wert 570 mg mit der Häufigkeit 12 am häufigsten vorkommt. **D = 570 mg**

Beispiel 10
Modalwert bei qualitativen Daten (Nominalwerte)
Nach der Differenzierung der Leukozyten einer Maus fragen wir, welche Zellform am Häufigsten vorkommt. Die Differenzierung ergab folgendes Ergebnis:

Zelltyp	10	20	30	40	50	60	70	80	90	100	Summe
Neutrophile	/	/	/	//	//	///	//		/	///	16
Basophile											0
Eosinophile	/										1
Monozyten		//	/	/						/	5
Lymphozyten	~~////~~ ///	~~////~~ //	~~////~~ ///	~~////~~ //	~~////~~ ///	~~////~~ //	~~////~~ ///	~~////~~ ~~////~~	~~////~~ ////	~~////~~ /	

Tabelle 8

Die hier untersuchte nominale Merkmalsausprägung, ist nicht messbar. Durch Vergleich ist festzustellen, ob die Ausprägung (z.B. Basophile Leucozyten) da ist und wenn ja, ihre Häufigkeit zu zählen. Rechnerisch lassen sich hier keine Mittelwerte bilden. Wir fragen nur: Welcher Zelltyp kommt am häufigsten vor: Mit 78 Werten bilden die Lymphozyten die am häufigsten vorkommende Zellform und damit den Modalwert.

D = Lymphozyten

4.2.5 Medianwert ($\tilde{x}$, M) (Median, Zentralwert)

gesprochen x Tilde

Eigenschaften	Stabil gegen Extremwerte
Anwenden	bei quantitativen Daten Intervalldaten Verhältnisdaten bei Ordinaldaten bei extrem kleinen Stichproben bei offenen Randklassen bei stark asymmetrischen Verteilungen bei multimodalen Verteilungen
Nicht anwenden	bei Nominaldaten wenn Extremwerte wichtig sind

Zur Einführung in den Medianwert wollen wir an einem Beispiel zeigen, dass hier das arithmetische Mittel als Mittelwert nicht geeignet ist.

Beispiel 11
Es liegt die Gehälterliste einer Gruppe von neun Personen vor, deren Bruttoeinkommen schon rangiert sind.

Monatsgehalt in €	810	810	810	850	850	850	900	900	4810	→ $\bar{x}$ = 1287,78
Ränge	1	2	3	4	5	6	7	8	9	

Das arithmetische Mittel ist mit 1287,78 € für die meisten Werte der Reihe nicht repräsentativ. Er liegt höher als 8 von 9 Werten und auch den 9. Wert repräsentier er nicht gut. Der Medianwert dagegen ist mit $\tilde{x}$ = 850 € zumindest für 8 von 9 Werten ein guter Repräsentant.

Zu Ermittlung des Medians werden die Daten nach $x_1 \leq x_2 \leq x_3 \leq x_4 \ldots \leq x_n$ rangiert. Alle Werte der gangierten Reihe werden beim kleinsten Wert beginnend, mit Rangzahlen von 1 bis n versehen. Als Medianwert gilt der mittlere Wert der geordneten Reihe. Er wird berechnet nach

Medianwert	$\tilde{x}$ = [(n + 1)/2] tes Merkmal der geordneten Reihe

$\tilde{x}$ = [(n + 1) / 2] ter Wert

$\tilde{x}$ = [10 / 2] ter Wert

$\tilde{x}$ = 5. Wert

$\tilde{x}$ **= 850**

Das mittlere Einkommen wird durch $\tilde{x}$ besser als durch $\bar{x}$ repräsentiert. Das arithmetische Mittel wird durch den Extremwert so sehr beeinflusst, dass er viel zu hoch liegt und somit die Mehrheit der

Werte nicht gut repräsentiert. Dass $\tilde{x}$ unempfindlich gegen Extremwerte (Ausreisser) ist, erkennen wir, wenn wir die 4810 € z.B. durch 9870 € ersetzen. $\tilde{x}$ ändert sich dadurch nicht.

Beispiel 12
$\tilde{x}$ **<u>Bei geradem n</u>**

Bei der Untersuchung von Zellen des menschlichen peripheren Blutes haben wir bei stabkernigen Granulozyten mikroskopisch deren Durchmesser in µm ermittelt. Einige Werte der Urliste liegen in der folgenden Tabelle rangiert vor.

µm geordnet	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.5	15.0
Ränge	1	2	3	4	5	6	7	8	9	10

<n/2 <n/2

$\tilde{x}$

n = 10

Tabelle 10

$\boxed{\tilde{x} = [(n + 1)/2] \text{ tes Merkmal der geordneten Reihe}}$

$\tilde{x}$ = [(10 + 1)/2] tes Merkmal

$\tilde{x}$ = 5.5. Merkmal

da es kein 5.5 tes Merkmal gibt, wird das arithmetische Mittel des 5. und 6. Merkmals gebildet.

$\tilde{x}$ = (12,0 + 12,5)/2

$\tilde{x}$ **= 12,25**

oder anders formuliert

$\boxed{\tilde{x} = 0{,}5\,(x_{\{n/2\}} + x_{\{n/2\}+1})}$ $x_{\{n/2\}}$ ist der Wert x bei n/2 = 10/2 = 5, also der Wert 12,0

$x_{\{n/2+1\}}$ ist der Wert x bei n/2 + 1 = 10/2 = 5 +1 = 6, also 12,5

$\tilde{x}$ = 0,5 (12,0 + 12,5)

$\tilde{x}$ **= 12,25**

Bei geradem n ist der Median nicht unbedingt ein realisierter Wert, d.h., er kommt nicht unbedingt in der Messwertreihe vor.

Es gilt: $\boxed{n/2 \leq \tilde{x} \leq n/2}$ Das bedeutet: 50% der Werte sind gleich oder kleiner als der Median und 50% der Werte sind größer oder gleich dem Median. Bei geradem n sind es jeweils genau 50%.

5 Werte (=50%) 12,25 5 Werte (50%)

Beispiel 13

$\tilde{x}$ **<u>Bei ungeradem n</u>**

Wir erweitern die Liste von Beispiel 12 um einen Wert.

geordnet	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.5	15.0	15.5
Ränge	1	2	3	4	5	6	7	8	9	10	11

< n/2 $\tilde{x}$ < n/2

n = 11

Tabelle 11

$\boxed{\tilde{x} = [(n + 1)/2] \text{ tes Merkmal der geordneten Reihe}}$

$\tilde{x}$ = 6. Merkmal

$\tilde{x}$ **<u>= 12,5</u>**

Bei ungeradem n ist der Median ein Wert, der in der Messdatenreihe realisiert ist (dort vorkommt).

Es gilt: $\boxed{n/2 \leq \tilde{x} \leq n/2}$

Das bedeutet: 50% der Werte sind gleich oder kleiner als der Median und 50% der Werte sind größer oder gleich dem Median.

5 Werte (<50%) 12,5 5 Werte (<50%)

Da bei ungeradem n der Median ein realisierter Wert ist, liegen unter und über ihm weniger als n/2, da der Medianwert ja selber in den 100% enthalten ist.

Beispiel 14

Wie wir vorgehen, wenn einzelne Messwerte mehrfach vorkommen, zeigt dieses Beispiel mit anderen Werten aus der Urliste von Beispiel 12.

geordnet	10.0	10.5	11.0	11.5	12.0	**12.5**	**12.5**	**12.5**	**12.5**	**13.0**	**13.0**	13.5	14.5	15.0
Ränge	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Tabelle 12

Jeder der mehrfach vorkommenden Werte wird in der geordneten Reihe nach $x_1 \leq x_2 \leq x_3 \leq \ldots. \leq x_n$ aufgeführt.

n= 14

$\tilde{x}$ = [(n + 1)/2] tes Merkmal der geordneten Reihe

$\tilde{x}$ = [(10 + 1)/2] tes Merkmal

$\tilde{x}$ = 7.5. Merkmal

$\tilde{x}$ = 12,5

Auch hier gilt $\boxed{n/2 \leq \tilde{x} \leq n/2}$

7 Werte ≤ 12,5 ≤ 7 Werte
= 50% = 50%

Beispiel 15
Dieses Beispiel soll den Zusammenhang zwischen arithmetischem Mittel, Modalwert und Medianwert zeigen. Während bei einer symmetrischen Verteilung die drei Mittelwerte identisch sind, zeigen asymmetrische Verteilungen bestimmte Muster in der Höhe der Mittelwerte. Dieses Beispiel zeigt dies für eine – fiktive - linksgipfelige Verteilung.

1	xi	0	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	90	100	110	120	130	140
2	Hi	0	5	15	35	65	77	80	76	68	58	50	42	35	30	25	21	15	10	7	5	3	2
3	ΣHi		5	20	55	120	197	277	353	421	479	529	571	606	636	661	682	697	707	714	719	722	724
4	xi*Hi	0	50	225	700	1625	2310	2800	3040	3060	2900	2750	2520	2275	2100	1875	1680	1350	1000	770	600	390	280

Tabelle 13

1. Berechnung des arithmetischen Mittels

$$\bar{x} = \frac{\sum_{i=1}^{k} (x_i * H_i)}{\sum_{i=1}^{k} H_i} = \frac{34300}{724} = 47,4$$

34300 ← (Summe der Werte in Zeile 4 der Tabelle)

724 ← (Summe der Werte in Zeile 2 der Tabelle)

$\bar{x}$ **= 47,4**

2. Ermittlung des Modalwertes

Der Modalwert ist der Wert, der am häufigsten vorkommt. Er ist über die Häufigkeiten in Zeile 2 der Tabelle abzulesen. $x_i = 35$ hat die Häufigkeit 80.

D = 35

3. Ermittlung des Medianwertes

$\tilde{x}$ = [(n + 1)/2] tes Merkmal der geordneten Reihe

$\tilde{x}$ = [(724+1)/2] tes Merkmal der geordneten Reihe

$\tilde{x}$ = 362,5. Merkmal der geordneten Reihe (siehe 3. Zeile der Tabelle)

$\tilde{x}$ **= 45**

Wenn die drei Mittelwerte in die Graphik (Abb. 3) eingezeichnet werden, dann erkennen wir die typischen Lagen der drei Werte bei einer linksgipfeligen Verteilung.

Allgemein gilt

Symmetrische Verteilung	D = $\tilde{x}$ = $\bar{\mathbf{x}}$
linksgipfelige Verteilung	D < $\tilde{x}$ < $\bar{\mathbf{x}}$
rechtsgipfelige Verteilung	D > $\tilde{x}$ > $\bar{x}$

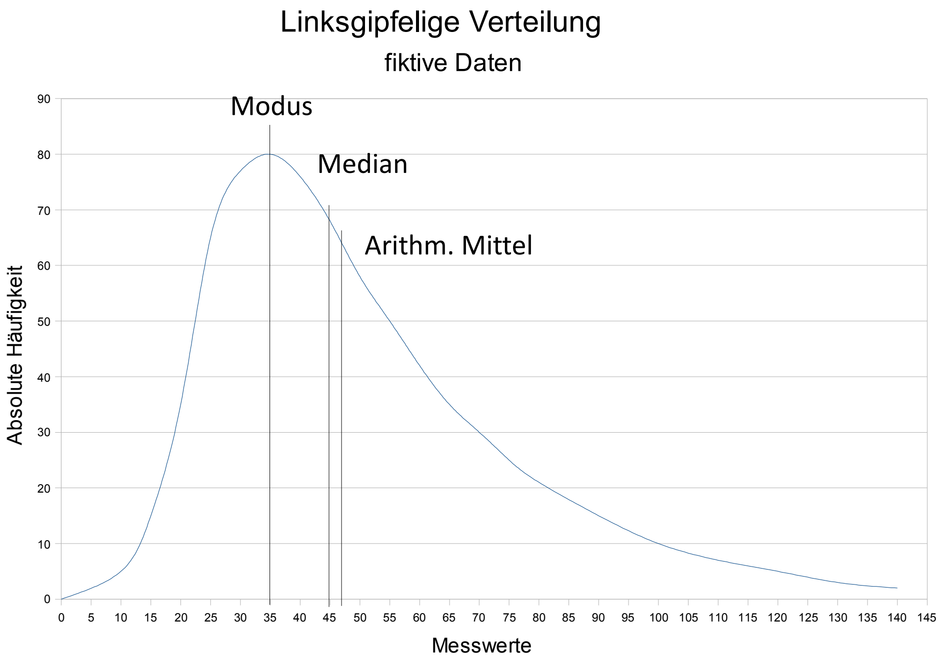


Abb.3

4.2.6 Quantile (Fraktile)

Diese Kennwerte werden in der Literatur gelegentlich auch bei den Kennwerten der Variation besprochen.

Eigenschaften	Stabil gegen Extremwerte
Anwenden	bei quantitativen Daten
Nicht anwenden	wenn Extremwerte wichtig sind.

Mit dem Medianwert haben wir den Kennwert kennengelernt, für den gilt

$$50\% \text{ aller Werte} \leq \tilde{x} \leq 50\% \text{ aller Werte}$$

das bedeutet, dass die Hälfte aller Werte kleiner oder gleich ist dem Median und höchstend die Hälfte alle Werte größer oder gleich dem Median ist. Der Medianwert teilt also die geordneten Daten in zwei gleich große Hälften von je 50%.

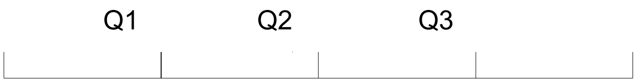
So wie wir mit dem Median die Grenze bestimmen können unterhalb derer 50% aller Werte liegen, können wir mit Kennwerten auch Grenzen bestimmen unter denen ein beliebiger Prozentsatz aller Werte liegt. Solche Kennwerte nennen wir Quantile, von denen es – je nach dem Prozentsatz –

verschiedene gibt. Ein Quantil ist also ein Punkt auf der Abszisse, für den gilt, dass ein gewählter Prozentsatz aller Daten kleiner oder gleich dem Quantil ist.

Häufig benutzte haben bestimmte Namen wie z.B.

Quartile	teilen Viertel ab
Dezile = Dezentile	teilen Zehntel ab
Zentile = Perzentile	teilen Hunderstel ab

Quartile
Quartile grenzen jeweils Viertel (Quartale) einer Datengruppe voneinander ab. Es gibt demnach also drei Trennpunkte, drei Quartile



Das 1. Quartil (Q1 oder $Q_{0,25}$) ist der Wert, unterhalb dessen 25% aller Werte liegen. Er wird errechnet nach

$$Q_{0,25} = (n+1) * 0{,}25.\ \text{Wert der geordneten Datenreihe}$$

$$\boxed{25\% \le Q_{0,25}}$$

Das 2. Quartil (Q2 oder $Q_{0,5}$) ist der Wert, unterhalb dessen 50% aller Werte liegen. Er wird errechnet nach

$$Q_{0,5} = (n+1) * 0{,}5.\ \text{Wert der geordneten Datenreihe, also der Medianwert}$$

$$\boxed{50\% \le Q_{0,5}}$$

Das 3. Quartil (Q3 oder $Q_{0,75}$) ist der Wert, unterhalb dessen 75% aller Werte liegen. Er wird errechnet nach

$$Q_{0,75} = (n+1) * 0{,}75.\ \text{Wert der geordneten Datenreihe}$$

$$\boxed{75\% \le Q_{0,75}}$$

Dezile
Entsprechend lassen sich Dezile berechnen, die jeweils Zehntel der Gesamtdaten abgrenzen. Es gibt neun Trennpunkte, neun Dezile. Sie werden wie folgt berechnet.

1. Dezil = $Q_{0,1} = (n+1) * 0{,}1$. Merkmal der geordneten Datenreihe
2. Dezil = $Q_{0,2} = (n+1) * 0{,}2$. Merkmal der geordneten Datenreihe
6. Dezil = $Q_{0,6} = (n+1) * 0{,}6$. Merkmal der geordneten Datenreihe
usw.

Zentile
Zentile teilen eine Datenreihe in hundertstel Teile. Es gibt 99 Zentile. Sie werden wie folgt berechnet.

1. Zentil = $Q_{0,01} = (n+1) * 0{,}01$. Merkmal der geordneten Datenreihe
65. Zentil = $Q_{0,65} = (n+1) * 0{,}65$. Merkmal der geordneten Datenreihe
80. Zentil = $Q_{0,80} = (n+1) * 0{,}8$. Merkmal der geordneten Datenreihe

Beispiel 16
Aus dem Beispiel mit den 257 Walnüssen (siehe Verteilungen) haben wir für dieses Beispiel 16 aus der Urliste die letzten 58 Werte aus Platzgründen gestrichen. Die verbleibenden 48 verschiedenen Werte, die sich auf 199 Einzeldaten verteilen, sind rangiert mit ihren Häufigkeiten in der folgenden Tabelle aufgelistet.

x_i	9.0	9.2	9.4	9.5	9.7	9.8	9.9	10.0	10.2	10.3
H_i	2	1	2	3	1	3	1	5	2	5
x_i	10.4	10.5	10.6	10.7	10.8	10.9	11.0	11.1	11.2	11.3
H_i	1	6	2	3	8	7	6	12	6	6
x_i	11.4	11.6	11.7	11.8	11.9	12.0	12.1	12.2	12.3	12.4
H_i	8	10	8	6	7	5	5	5	6	2
x_i	12.5	12.6	12.7	12.8	12.9	13.0	13.1	13.2	13.4	13.5
H_i	6	3	5	5	3	5	3	4	2	2
x_i	13.6	13.7	13.8	14.0	14.1	14.4	14.5	14.8	15	
H_i	2	3	2	3	2	2	1	1	1	

Tabelle 14

Wir wollen an diesen Zahlen die Berechnung folgender Quantile Durchführen:
$Q_{0,25}$; $Q_{0,05}$,; $Q_{0,3}$; $Q_{0,68}$; $Q_{0,75}$.

$Q_{O,25}$ ist der Wert, unterhalb dessen 25% aller Werte liegen.
$Q_{O,25}$ = (n + 1) * 0,25. Wert der geordneten Datenreihe
$Q_{O,25}$ = (199 + 1) * 0,25. Wert der geordneten Datenreihe
$Q_{O,25}$ = 200 * 0,25. Wert der geordneten Datenreihe
$Q_{O,25}$ = 50. Wert der geordneten Datenreihe

Wenn alle 199 Werte der geordneten Urliste vorliegen würden, dann bräuchten wir nur von x_1 bis x_{50} gehen und hätten dann den 50. Wert. Unsere Liste ist aber komprimiert und enthält nur jeden Messwert mit seiner Häufigkeit. Um hier zum 50. Wert zu gelangen müssen wir Häufigkeiten in den Zeilen H_i bis zum 50. Wert aufsummieren und finden als 50. x_i-Wert 10,9.

$Q_{O,25}$ = 10,9

49 Werte = 24,6% aller Werte liegen unter $Q_{O,25}$ = 10,9

Es gilt also $\boxed{25\% \leq Q_{0,25}}$

$Q_{O,05}$ = (199 + 1) * 0,05. Wert der geordneten Datenreihe
$Q_{O,05}$ = 200 * 0,05. Wert der geordneten Datenreihe
$Q_{O,05}$ = 10. Wert der geordneten Datenreihe
$Q_{O,05}$ = 9,8

9 Werte = 4,5% aller Werte liegen unter $Q_{O,05}$ = 9,8

Es gilt also $\boxed{5\% \leq Q_{0,25}}$

$Q_{O,3}$ = (199 + 1) * 0,3. Wert der geordneten Datenreihe
$Q_{O,3}$ = 60. Wert der geordneten Datenreihe
$Q_{O,3}$ = 11,1

59 Werte = 29,6% aller Werte liegen unter $Q_{O,3}$ = 11,1

Es gilt also $\boxed{30\% \leq Q_{0,25}}$

$Q_{O,68}$ = (199 + 1) * 0,68. Wert der geordneten Datenreihe
$Q_{O,68}$ = 136. Wert der geordneten Datenreihe
$Q_{O,68}$ = 12,2

135 Werte = 67,8% aller Werte liegen unter $Q_{O,68}$ = 12,2

Es gilt also $\boxed{68\% \leq Q_{0,25}}$

$Q_{O,75}$ = (199 + 1) * 0,75. Wert der geordneten Datenreihe
$Q_{O,75}$ = 150. Wert der geordneten Datenreihe
$Q_{O,75}$ = 12,5

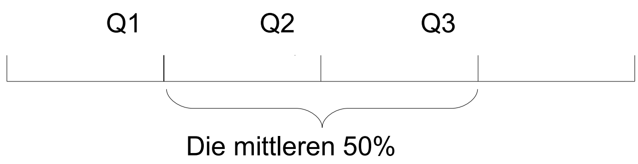
149 Werte = 74,9% aller Werte liegen unter $Q_{O,75}$ = 12,5

Es gilt also $\boxed{75\% \leq Q_{0,25}}$

Häftespielraum, Quartilabstand (QA)

Hier liegt der Grund, weswegen die Quantile gelegentlich auch bei den Kennwerten der Variation genannt werden.

Der Quartilabstand ist der Bereich auf der Skala, der zwischen Q1 und Q3 liegt. Es ist der Bereich der Daten, in dem die mittleren 50% aller Werte liegen.



Er wird berechnet nach

$$\boxed{\begin{array}{l} QA = |Q3 - Q1| \\ QA = |Q_{0,75} - Q_{0,25}| \end{array}}$$

Für die Daten von Beispiel 16 gilt

$$QA = |Q_{0,75} - Q_{0,25}|$$
$$QA = 12,5 - 10,9$$

Zwischen 10,9 und 12,5 liegen, wenn Sie mal nachzählen, 100 Werte = 50,25%.

Wir werden im nächsten Kapitel bei der Kennwerten der Variation näher auf den Quartilenabstand eingehen.

4.3 Übungen

Übung 1
Berechnen Sie das gemeinsame arithmetische Mittel für alle Messwerte der Stichproben 1 - 4.

Stichprobe	n	$\bar{x}$
1	7	0.136
2	14	0.179
3	19	0.094
4	4	0.125

Übung 2
Bei einer wachsenden Bakterienkultur wurden über 5 Stunden Keimzahlbestimmungen durchgeführt. Die Ergebnisse waren.

Anfang der 1. Stunde:	10	* 10^3 K/mL
Anfang der 2. Stunde:	25	* 10^3 K/mL
Anfang der 3. Stunde:	75	* 10^3 K/mL
Anfang der 4. Stunde:	0,275	* 10^6 K/mL
Anfang der 5. Stunde:	1,072	* 10^6 K/mL
Ende der 5.Stunde :	3,859	* 10^6 K/mL

Berechnen Sie das geometrische Mittel als mittleren Vermehrungsfaktor. Weisen Sie nach, dass dieser Wert richtig ist und nicht das arithmetische Mittel.

Übung 3
Wir haben bei Ratten (250 g bis 270 g, männlich, Wistar) die Nierenfunktion untersucht und dabei die endogene Kreatinin-Clearance bestimmt. Die Ergebnisse sind:

Tier Nr.	x_i mL/min	Tier Nr.	x_i mL/min	Tier Nr.	x_i mL/min
1	0,15	11	0,32	21	0,09
2	0,36	12	0,34	22	0,65
3	0,44	13	0,21	23	0,44
4	0,22	14	0,33	24	0,53
5	0,23	15	0,16	25	0,15
6	0,64	16	0,55	26	0,12
7	0,09	17	0,41	27	0,51
8	0,37	18	0,39	28	0,29
9	0,56	19	0,19	29	0,23
10	0,35	20	0,32	30	0,32

Ermitteln Sie arithmetisches Mittel, Modus und Median.

Übung 4

Bei einem toxikologischen Langzeitversuch wurden 15 Monate nach Versuchsbeginn die Überlebenszeiten von 12 Ratten wie folgt festgestellt.

Tier	überlebte Tage
1	285 d
2	320 d
3	375 d
4	275 d
5	überlebt
6	110 d
7	überlebt
8	300 d
9	überlebt
10	295 d
11	395 d
12	400 d

Berechnen Sie das harmonische Mittel als mittlere Überlebenszeit. Die überlebenden Tiere sind irgendwann nach 15 Monaten gestorben. Deren Todeszeitpunkt stand aber nicht mit dem Versuch im Zusammenhang. Diese Überlebenszeiten werden gleich ∞ gesetzt. Beachten Sie $1/\infty = 0$

Übung 5

Im anatomischen Praktikum haben wir die Schilddrüsen von 30 männlichen Ratten (250 g bis 270 g, Wistar) in mg gewogen und als Urliste notiert:

x_i	9.8	7.9	6.9	11.0	12.0	12.2	13.6	14.6	9.8	7.5
x_i	10.3	15.0	9.6	10.0	10.5	11.7	8.5	7.9	9.8	6.9
x_i	11.0	14.5	10.5	8.1	8.9	9.3	6.8	10.0	10.3	12.1

Bestimmung Sie Modalwert, Medianwert und das arithmetische Mittel.

Übung 6

Ermitteln Sie den Hälftespielraum ($|Q1 - Q3|$) für die Daten von Übung 5

Lösungen

Übung1:$\bar{x}$ = 0,1305. Übung 2: $\bar{x}$ = 3,34; $\bar{x}G$ = 3,30. Übung 3: $\bar{x}$ = 0,332; Modus = 0,32; Median = 0,325. Mit $\bar{x}$ = 3,34 erreichen wir am Ende der 5. Stunde 4,156 * 10^6 Zellen/mL. Mit $\bar{x}_G$ = 3,30 erreichen wir am Ende der 5. Stunde 3,896 * 10^6 Zellen/mL. Die Gründe für die Abweichung 3,896 * 10^6 Zelle/mL - 3,859 * 10^6 Zellen/mL: Die stündlichen Faktoren habe ich auf eine Nachkommastelle angegeben (2,5;3,0;3,7;3,9;3,6). Das geometrische Mittel der gerundeten Faktoren habe ich mit zwei Nachkommastellen angegeben. Die Rundungs“fehler“ gehen multiplikativ in das Ergebnis ein. Übung 4: $\bar{x}_H$ = 355 d. Übung 5: Arithmetisches Mittel: 10,23 mg,Medianwert 10 mg, Modalwert 9,8 mg. Übung 6: Q1 = 7,75; Wert = 8,5; Q3 = 23,25; Wert = 17,7. Zwischen Q1 und Q3 liegen 16 Werte = 53%.