

Einführung in die beschreibende Statistik

Beispiel einer Datenerhebung mit Begriffserklärungen (Vokabel)

Was ist an der Behauptung dran „Unsere Jugend wird immer dicker und träger“?
Um eine Behauptung dieser Art zu überprüfen kann man nicht alle Jugendlichen befragen, sondern nur einen Teil der Gesamtheit.

Stichprobe: Wird der Teil einer Gesamtheit befragt, dann spricht man bei der Datenerhebung von einer Stichprobe.

Urliste: Das Ergebnis der Stichprobe wird in einer Urliste festgehalten.

Beispiel einer Urliste:

Schüler Nr.	Geschlecht		Körpergröße in cm	Gewicht in kg	Raucher		Sportart
	w	m			ja	nein	
01		x	160	52		x	Handball
02	x		172	67	x		Fußball
03	x		180	60	x		Golf
...
xx		x	170	65		x	Judo

(Eine entsprechende Urliste ist in der jeweiligen Klasse anonym zu erstellen)

Rohdaten: Sind alle in der Urliste enthaltenen Daten.

Erhebungsumfang: Ist die Anzahl der untersuchten Objekte (hier Schüler)
Werden z.B. 27 Schüler befragt, so sagt man,
“Die Anzahl der Merkmalsträger ($n = 27$) bildet den Erhebungsumfang“.

Merkmale: Sind die Eigenschaften der Objekte.
(z.B. Geschlecht, Körpergröße, Gewicht, Raucher, Sportart, ...)

Merkmalsausprägung x_i : Ein Merkmal kann in verschiedenen Ausprägungen vorkommen. (z.B. Geschlecht m oder w)

Beispiele für Merkmale und deren Ausprägungen:

Merkmal	Merkmalsausprägung x_i
Geschlecht	männlich, weiblich
Körpergröße	160 cm, 182 cm, 154 cm, 163 cm, ...
Körpergewicht	52 kg, 81 kg, 71 kg,
Raucher	ja, nein
Sportart	Judo, Tischtennis, Fußball, ...

Bemerkung: Durch eine Erhebung soll festgestellt werden, wie die verschiedenen Ausprägungen eines Merkmals in einer Stichprobe verteilt sind.
Vor der Erhebung sind einige Fragen zu klären, z.B.
Wonach soll gefragt werden?

Welche Antworten sind möglich?
Welchen Umfang soll die Stichprobe haben?

Aufbereiten der Daten und ihre Darstellung.

Ermittlung der Häufigkeit bestimmter Merkmalsausprägungen

In einer **Strichliste** werden Merkmalsträger mit gleichen Merkmalsausprägungen für ein bestimmtes Merkmal zusammengefasst.

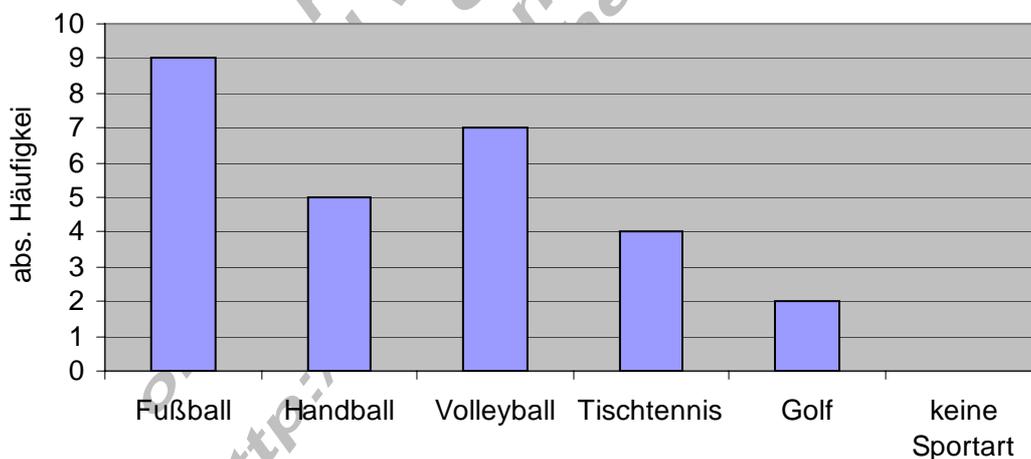
Bearbeitung des Merkmals *Sportart* (eventuell in Klassen zusammenfassen)

Merkmalsausprägung		absolute Häufigkeit
Fußball	/// IIII	9
Handball	///	5
Volleyball	/// II	7
Tischtennis	IIII	4
Golf	II	2
keine Sportart		0

Der Merkmalsausprägung z.B. Fußball wird die absolute Häufigkeit 9 zugeordnet.

Grafische Darstellung durch ein **Säulendiagramm**.

Säulendiagramm



Berechnung des Body - Mass - Index (BMI):

$$\text{BMI} = \frac{\text{Gewicht in kg}}{(\text{Körpergröße in m})^2}$$

Wertung:

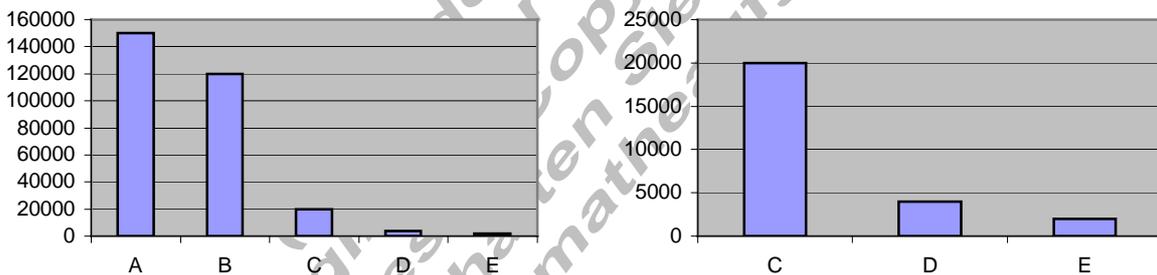
I) $0 < x < 20$ Untergewicht	II) $20 \leq x \leq 25$ normal
III) $25 < x \leq 30$ Übergewicht	IV) $30 < x \leq 35$ moderate Fettsucht
V) $35 < x$ morbide Fettsucht	

Auszug aus der Urliste mit Berechnung des BMI

Schüler Nr.	Geschlecht		Körpergröße in cm	Gewicht in kg	BMI	Wertung
	m	w				
01		x	160	52	20,3	normal
02	x		172	67	22,6	normal
03	x		180	60	18,5	Untergewicht
04		x	167	55	19,7	Untergewicht
05		x	178	63	19,9	Untergewicht
...

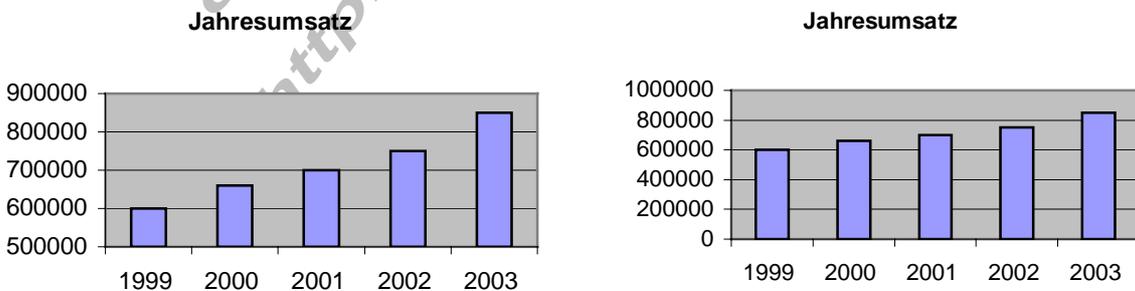
Einfluss der Darstellung auf die Interpretation der Daten

Beispiel 1:



Durch Weglassen der Werte für A und B und durch eine andere Skalierung erscheint im zweiten Diagramm der Wert für C ziemlich gigantisch.

Beispiel 2



Durch die unterschiedliche Achsenteilung erscheint die Umsatzentwicklung im linken Diagramm wesentlich günstiger zu sein als im rechten.

Lagemaße

Zur Interpretation einer Beobachtungsreihe kann man neben der grafischen Darstellung weitere charakteristische Größen heranziehen.

Mittelwert und Median

Arithmetisches Mittel einer Datenreihe

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$$

n : Anzahl der Beobachtungswerte x_i x_i : i -ter Beobachtungswert $i \in \mathbb{N}$

Beispiel:

Bestimmen Sie aus der Liste einer Schülerbefragung die durchschnittliche Körpergröße aller befragten Schüler.

Körpergröße

x_i : 164 ; 162 ; 168 ; 190 ; 166 ; 166 ; 162 ; 160 ; 150 ; 175 ; 180 ; 160 ; 168 ; 175 ; 168

Anzahl der Beobachtungswerte (Schüler) $n = 15$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{15} \sum_{i=1}^{15} x_i$$

$$= \frac{1}{15} (164 + 162 + 168 + 190 + 166 + 166 + 162 + 160 + 150 + 175 + 180 + 160 + 168 + 175 + 168)$$

$$= \frac{2514}{15} = \underline{\underline{167,6}} \text{ (Durchschnittsgröße)}$$

Weitere Beispiele für Mittelwerte:

Durchschnittsabiturnote: 1,8

Durchschnittsgewicht aller Schüler einer Klasse: 62,3 kg

Der Median (Zentralwert einer Datenreihe)

Der Median x_{Med} ist derjenige Wert (Merkmalsausprägung), der in der Mitte steht, wenn alle Beobachtungswerte x_i der Größe nach geordnet sind.

Wir ordnen alle Werte aus unserem Beispiel der Größe nach und bestimmen die Mitte.

Körpergröße geordnet

x_i : 150 ; 160 ; 160 ; 162 ; 162 ; 164 ; 166 ; **166** ; 168 ; 168 ; 168 ; 175 ; 175 ; 180 ; 190

Der 8. Wert, also $x_8 = 166$ bildet den Median.

Wie verändern sich Mittelwert und Median, wenn der größte Schüler die Klasse verlässt und für ihn eine kleine Schülerin mit der Körpergröße 150 dazu kommt?

Körpergröße geordnet

x_i : 150 : 150 ; 160 ; 160 ; 162 ; 162 ; 164 ; **166** ; 166 ; 168 ; 168 ; 168 ; 175 ; 175 ; 180

Der Median, also $x_8 = 166$ bleibt unverändert.

$$\text{Mittelwert: } \bar{x} = \frac{1}{15} \sum_{i=1}^{15} x_i$$

$$= \frac{1}{15} (164 + 162 + 168 + 150 + 166 + 166 + 162 + 160 + 150 + 175 + 180 + 160 + 168 + 175 + 168)$$

$$= \frac{2474}{15} = \underline{\underline{164,9}}$$

Der Mittelwert wird nach unten gezogen.

Wie verändert sich der Median, wenn ein Schüler mit der Körpergröße 180 dazu kommt?

Körpergröße geordnet

x_i : 150 : 150 ; 160 ; 160 ; 162 ; 162 ; 164 ; **166** ; **166** ; 168 ; 168 ; 168 ; 175 ; 175 ; 180 ; 180

Der Median liegt zwischen den Werten von x_8 und x_9 , bleibt also bei 166

Allgemeine Rechenvorschrift zur Berechnung des Medians

Ist n ist die Anzahl der Beobachtungswerte x_i , dann gilt:

$$n \text{ ungerade} \Rightarrow x_{\text{Med}} = x_{\frac{n+1}{2}} \quad n \text{ gerade} \Rightarrow x_{\text{Med}} = \frac{1}{2} \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right)$$

Ergänzungen zum Median

Beispiel:

Ein Baupersonal mit 9 Personen hat folgende monatliche Einkünfte in Euro.

1160 | 1050 | 980 | 1200 | 970 | 1800 | 6600 | 1180 | 1090

arithmetisches Mittel: $\bar{x} \approx 1781,11 \text{ €}$

Dieser Durchschnitt liefert ein falsches Bild, weil die Mehrzahl (7 von 9 Personen) höchstens 1200 € verdient.

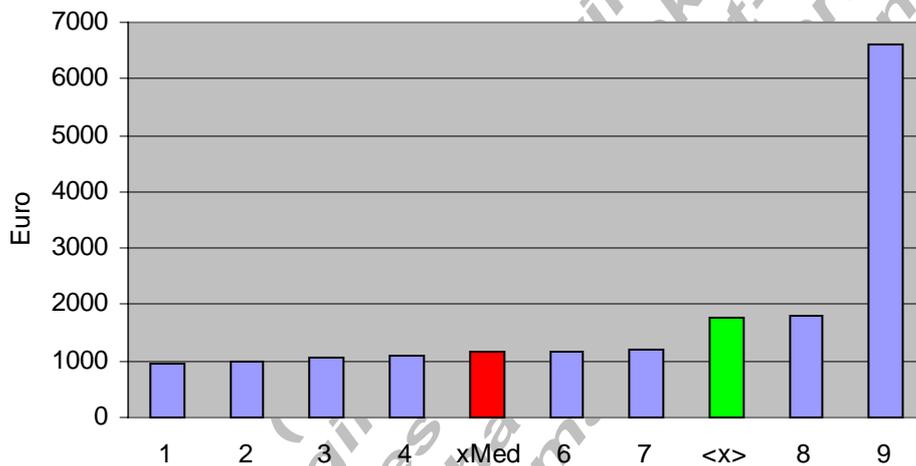
Der Wert 6600 zieht den Mittelwert nach oben.

Wir berechnen den Median

970 | 980 | 1050 | 1090 | **1160** | 1180 | 1200 | 1800 | 6600 $n = 9$ ungerade

$$\Rightarrow x_{\text{Med}} = x_{\frac{n+1}{2}} = x_{\frac{9+1}{2}} = x_5 = 1160$$

Der Median beschreibt die Verteilung besser als der Mittelwert. Man nennt ihn auch Zentralwert.



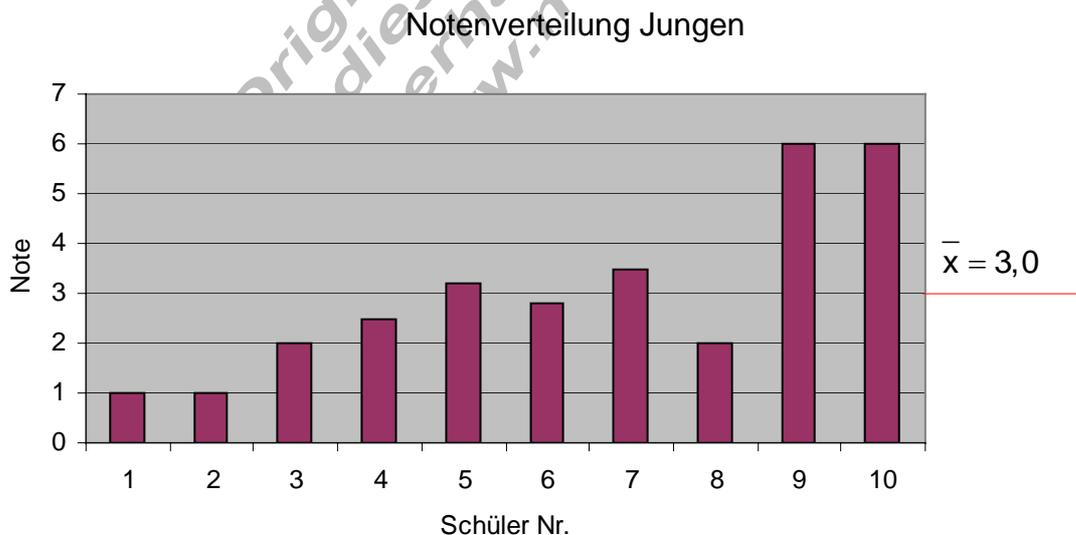
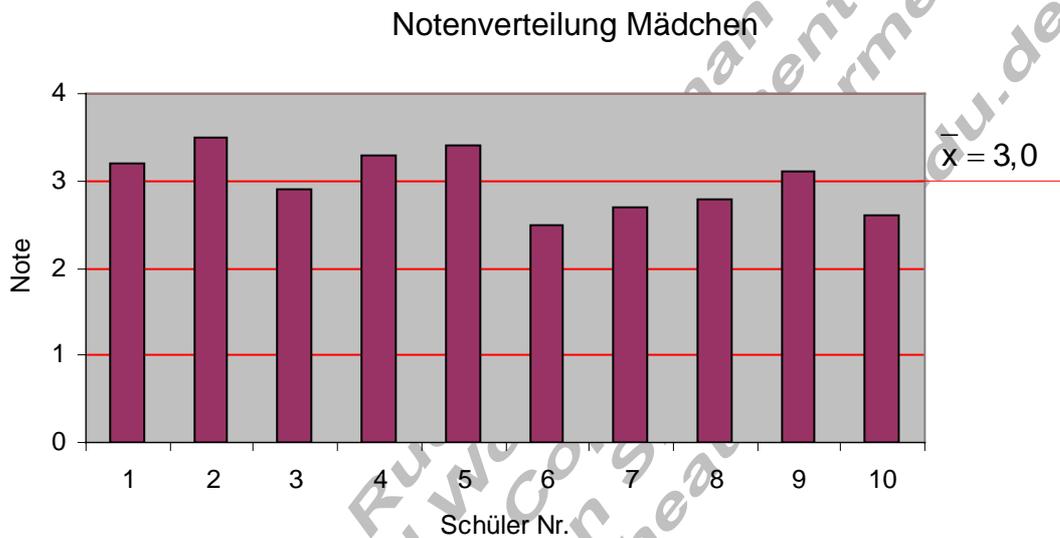
Ausreißer haben auf den Median keinen Einfluss.

Spannweite, Median, Varianz und Standardabweichung

Streuung um den Mittelwert.

In den folgenden Säulendiagrammen ist die Notenverteilung zweier Schülergruppen (Mädchen, Jungen) dargestellt, deren Mittelwert gleich ist.

Schüler Nr.	1	2	3	4	5	6	7	8	9	10	
Note Mädchen	3,2	3,5	2,9	3,3	3,4	2,5	2,7	2,8	3,1	2,6	$\bar{x} = 3,0$
Note Jungen	1,0	1,0	2,0	2,5	3,2	2,8	3,5	2,0	6,0	6,0	$\bar{x} = 3,0$



Bei den Mädchen liegen die Noten alle sehr nahe am Mittelwert
Sie streuen wenig um den Mittelwert.

Bei den Jungen sind die Abweichungen vom Mittelwert sehr groß.

Sie streuen stark um den Mittelwert.
Die Statistik bietet Möglichkeiten, die Streuung näher zu untersuchen.

(C) Rudolf Brinkman
Original Word- Dokumente
ohne diesen Copyright- Vermerk
erhalten Sie unter:
<http://www.matheaufgaben-du.de>

Die Spannweite.

Berechnet man den Unterschied zwischen dem größten und kleinsten Beobachtungswert, so erhält man die Spannweite.

Sie ist ein Maß für die Breite des Streubereichs einer Häufigkeitsverteilung.

Die Spannweite

Spannweite = größter Beobachtungswert - kleinster Beobachtungswert

$$R = x_{\max} - x_{\min}$$

Beispiel:

Schüler Nr.	1	2	3	4	5	6	7	8	9	10	
Note Mädchen	3,2	3,5	2,9	3,3	3,4	2,5	2,7	2,8	3,1	2,6	$\bar{x} = 3,0$
Note Jungen	1,0	1,0	2,0	2,5	3,2	2,8	3,5	2,0	6,0	6,0	$\bar{x} = 3,0$

Spannweite Mädchen: $R_M = 3,5 - 2,5 = \underline{1}$

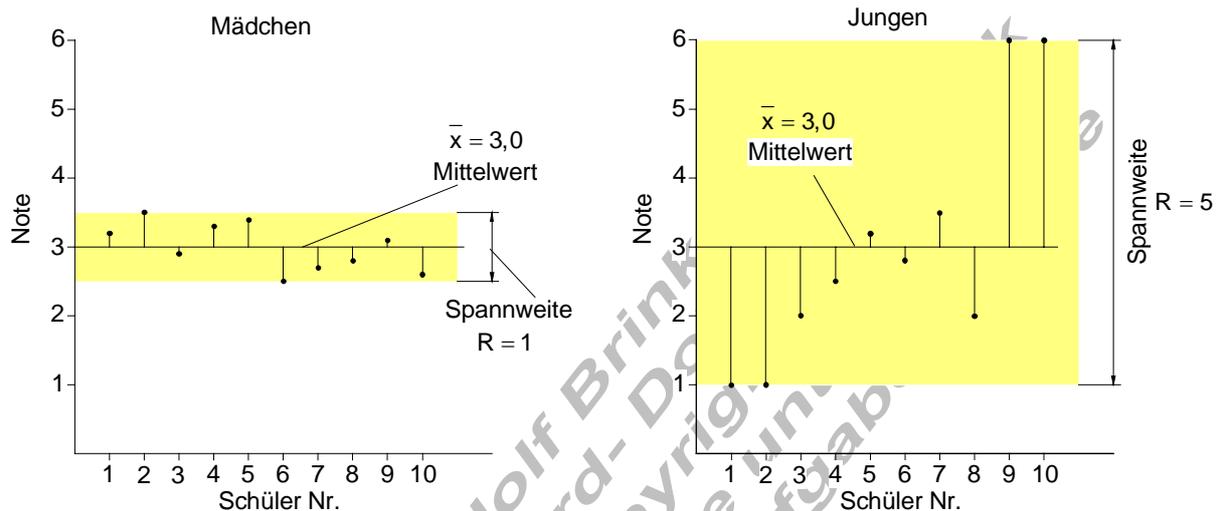
Spannweite Jungen: $R_J = 6,0 - 1,0 = \underline{\underline{5}}$

Varianz und Standardabweichung.

Wir betrachten noch mal die Notenverteilung von Mädchen und Jungen aus dem vorigen Beispiel.

Schüler Nr.	1	2	3	4	5	6	7	8	9	10	
Note Mädchen	3,2	3,5	2,9	3,3	3,4	2,5	2,7	2,8	3,1	2,6	$\bar{x} = 3,0$
Note Jungen	1,0	1,0	2,0	2,5	3,2	2,8	3,5	2,0	6,0	6,0	$\bar{x} = 3,0$

Der Mittelwert ist in beiden Fällen gleich, die Streuung um diesen ist unterschiedlich.



Abweichung

Die beiden Diagramme veranschaulichen die Abweichungen vom Mittelwert.

Abweichung: $x_i - \bar{x}$

In der beschreibenden Statistik berechnet man das arithmetische Mittel der Abweichungsquadrate und nennt dieses die **Varianz**.

Varianz einer Datenreihe

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

n: Anzahl der Beobachtungswerte, x_i : i-ter Beobachtungswert, \bar{x} : Mittelwert

Für unser Beispiel gilt:

Mädchen				Jungen			
i	x_i	\bar{x}	$(x_i - \bar{x})^2$	i	x_i	\bar{x}	$(x_i - \bar{x})^2$
1	3,2	3	0,04	1	1,0	3	4,0
2	3,5	3	0,25	2	1,0	3	4,0
3	2,9	3	0,01	3	2,0	3	1,0
4	3,3	3	0,09	4	2,5	3	0,25
5	3,4	3	0,16	5	3,2	3	0,04
6	2,5	3	0,25	6	2,8	3	0,04
7	2,7	3	0,09	7	3,5	3	0,25
8	2,8	3	0,04	8	2,0	3	1,0
9	3,1	3	0,01	9	6,0	3	9,0
10	2,6	3	0,16	10	6,0	3	9,0
Σ	30		1,10	Σ	30		28,58

Varianz Mädchen: $s_M^2 = \frac{1}{10} \cdot 1,1 = \underline{\underline{0,11}}$ Varianz Jungen: $s_J^2 = \frac{1}{10} \cdot 28,58 = \underline{\underline{2,858}}$

Viele Daten sind mit Einheiten behaftet, z.B. Meter (m) oder kg.

Die Einheit für die Varianz wäre in diesen Fällen m^2 bzw. $(kg)^2$.

Um wieder auf die ursprüngliche Einheit zu kommen, zieht man die Wurzel aus der Varianz. Dieser Wert wird **Standardabweichung** genannt.

Standardabweichung: $s = \sqrt{s^2} = \sqrt{\text{Varianz}}$
--

Zur praktischen Berechnung fertigt man wie oben gezeigt eine entsprechende Tabelle an. Sie dient auch zur Kontrolle der Daten.

Die Standardabweichung ist ein Maß dafür, wie hoch die Aussagekraft des Mittelwertes ist.

Eine kleine Standardabweichung bedeutet, alle Beobachtungswerte liegen nahe am Mittelwert.

Eine große Standardabweichung bedeutet, die Beobachtungswerte sind weit um den Mittelwert gestreut.

