Technische Hochschule Ingolstadt

Statistik

Themengebiet: Deskriptive Statistik

VE 03: Lage- und Streuungsmaße

Wintersemester 2024/25

Dozent: Prof. Dr. Sören Gröttrup

Vorlage von Prof. Dr. Max Krüger

Fakultät Informatik, Technische Hochschule Ingolstadt (THI)

Lerninhalte VE 03

Thema: Lage- und Streuungsmaße

- Mittelwert und gewichteter Mittelwert
- Median
- Quantil
- Modalwert
- Geometrisches und harmonisches Mittel
- Spannweite
- Mittlere Abweichung vom Mittelwert/Median
- Varianz und Standardabweichung
- Quartilsabstand

Inhaltsverzeichnis VE 03

VE 03: Lage- und Streuungsmaße

Beispiel: Firma MitarbeiterFroh

In unserer Finanzdienstleistungsfirma **MitarbeiterFroh** *verdienen die Mitarbeiter durchschnittlich* ***über 11.000 €*** *pro Monat bei unter 40 h Arbeitszeit pro Woche!*

Steigen auch Sie als Finanzdienstleister bei uns ein!

<u>Aufgabe</u>:

Berechnen Sie den

- *durchschnittlichen monatlichen Verdienst und die*
- *durchschnittliche wöchentliche Arbeitszeit,*

der Mitarbeiter für eine Firma in der 990 Mitarbeiter eine 39 h-Woche haben und dabei 1500 € im Monat verdienen und weitere 10 Topmitarbeiter bei einer 70 h- Woche ein monatliches Gehalt von 1.000.000 € erhalten.

Beispielhafte Fragestellungen

Beim Vergleich von Stichproben bzw. den ihnen zugrundeliegenden Grundgesamtheiten ergeben sich oftmals Fragen der folgenden Art:

- Wo liegt das Zentrum?
- Wie stark streuen die Werte?
- Ist die Verteilung symmetrisch oder schief?
- Gibt es Ausreißer?
- etc. .

Begriffsbildung zu Lage- und Streuungsmaßen

Maßzahlen beschreiben die Eigenschaften einer Stichprobe oder Grundgesamtheit in komprimierter Form durch numerische Werte formal.

Lagemaße beschreiben die Lage einer Stichprobe oder Grundgesamtheit. Welches Lagemaß bei einer bestimmten Fragestellung sinnvoll ist, hängt vom Kontext, von der Datensituation und vom Skalenrangniveau des Merkmals ab.

Streuungsmaße beschreiben die Streuung, d.h. die Abstandseigenschaften der einzelnen Stichprobenwerte oder Grundgesamtheitswerde von den zugehörigen Mittelwerten.

Inhaltsverzeichnis VE 03

VE 03: Lage- und Streuungsmaße

Empirischer Mittelwert (Lagemaß)

Empirischer Mittelwert (arithmetisches Mittel)

Sei $x = (x_1, \ldots, x_n)$ eine Stichprobe vom Umfang n, dann heißt

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

der (**empirische**) **Mittelwert** (bzw. **arithmetisches Mittel**) der Stichprobe x.

Berechnung mit Häufigkeiten

Kommen in der Stichprobe $x = (x_1, \ldots, x_n)$ genau die Werte $y_1, \ldots, y_k$ vor mit den relativen Häufigkeiten $h(y_1), \ldots, h(y_k)$, so kann der (empirische) Mittelwert der Stichprobe auch wie folgt berechnet werden:

$$\bar{x} = y_1 \cdot h(y_1) + \cdots + y_k \cdot h(y_k) = \sum_{i=1}^{k} y_i \cdot h(y_i)$$

Beispiel empirischer Mittelwert

Für ein neues Automodell der gehobenen Klasse ermittelt der ADAC jeweils den genauen Benzinverbrauch auf 100 km bei unterschiedlichen Streckenprofilen:

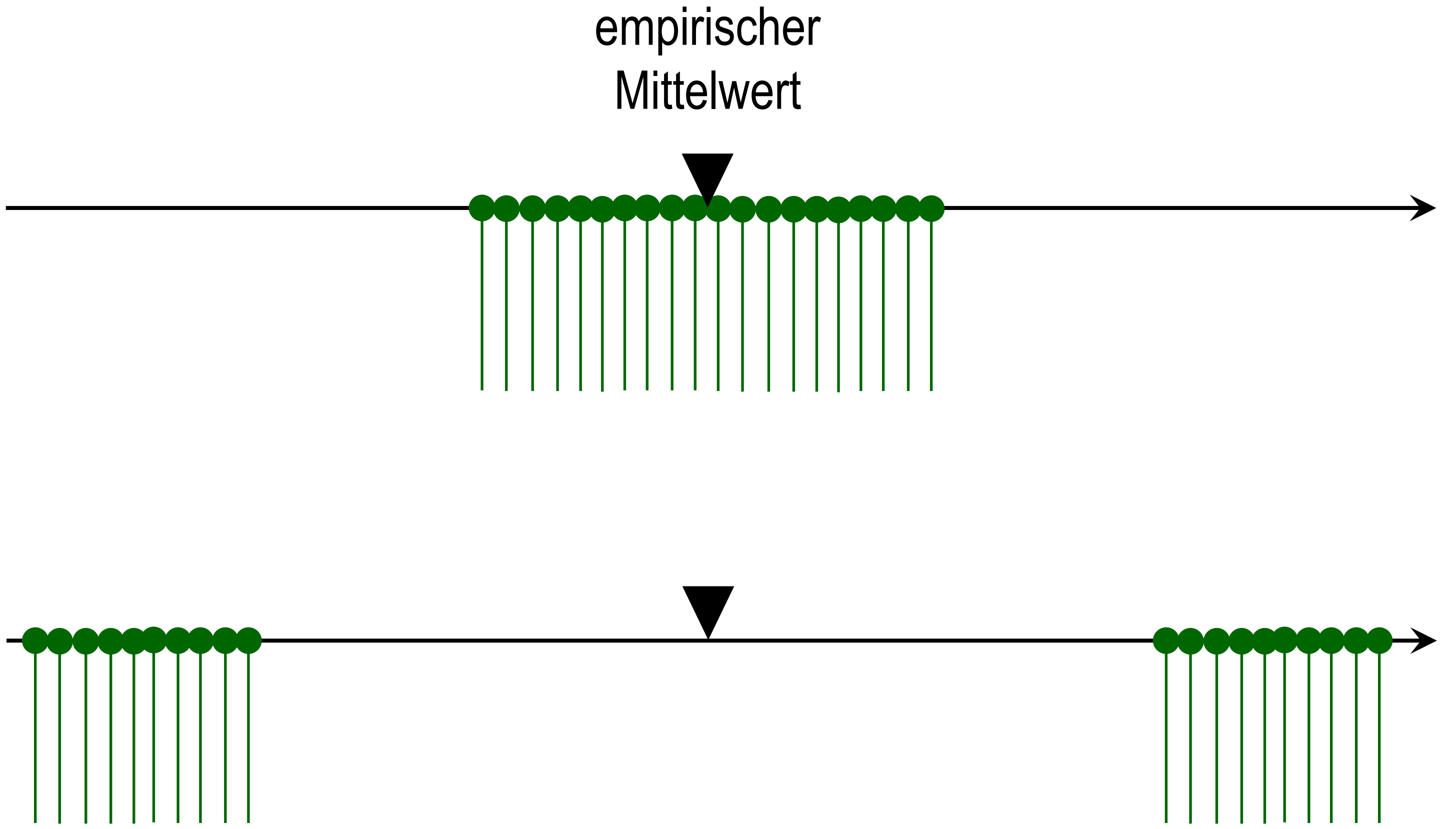
Streckenprofil	1	2	3	4	5	6	7	8
Verbrauch in l/100 km	7.3	9.5	8.4	6.9	10.3	10.2	8.5	7.9

Berechnung des mittleren Verbrauchs (empirischer Mittelwert):

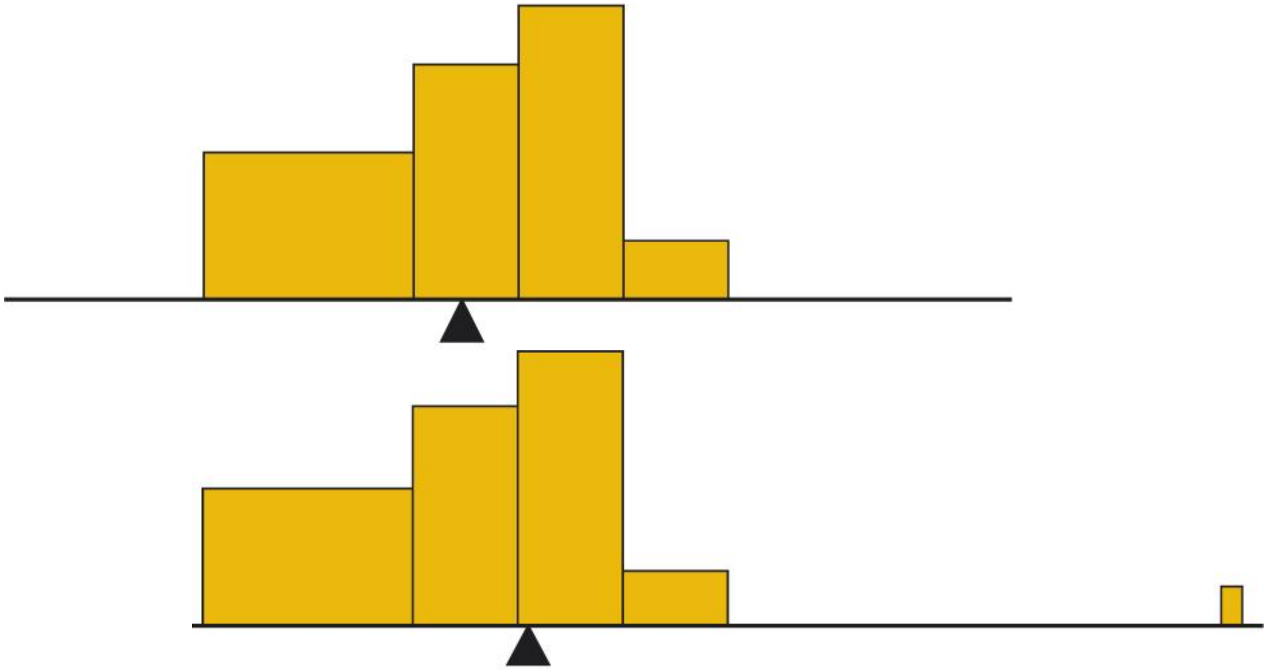
$$(7.3+9.5+8.4+6.9+10.3+10.2+8.5+7.9)\ /\ 8=8.625,$$

d.h. 8.625 Liter pro 100 km.

Problem empirischer Mittelwert (1)



Problem empirischer Mittelwert (2)



Aus: Arens et al., *Mathematik*, ISBN: 978-3-8274-1758-9

Eigenschaft:

Der empirische Mittelwert reagiert empfindlich auf (einzelne) extreme Werte oder Ausreißer in den Daten. (Hebelwirkung!)

Gewichteter empirischer Mittelwert (Lagemaß)

Gewichtete (**empirische**) **Mittelwert**

Sei $x = (x_1, \ldots, x_n)$ eine Stichprobe, in der die Stichprobenelemente unterschiedliche Gewichte $w_1, \ldots, w_n$ haben, dann heißt der Quotient

$$\bar{x}_w = \frac{w_1 \cdot x_1 + \cdots + w_n \cdot x_n}{\sum_{i=1}^{n} w_i}$$

der **gewichtete** (**empirische**) **Mittelwert** (bzw. **gewichtetes arithmetisches Mittel**) der Stichprobe x.

Anmerkung: Der gewichtete empirische Mittelwert wird angewandt, wenn die auftretenden Messwerte unterschiedliches Gewicht haben.

Beispiel gewichteter empirischer Mittelwert

Die Überlebensrate neugeborener Wildtiere ist saisonal stark schwankend.

Zeitraum	Dauer	Überlebensrate
Dezember-Januar	2	20 %
Februar-Juni	5	75 %
Juli-September	3	40 %
Oktober-November	2	35 %

Die durchschnittliche Überlebensrate ist (mit der Dauer als Gewicht) damit

$$(2*20\ \% + 5*75\% + 3*40\ \% + 2*35\%)/(2+5+3+2) = 50{,}41666\ldots\%$$

Das nicht gewichtete empirische Mittel liefert den falschen Wert 42,5 %.

Empirischer Median (Lagemaß)

Median

Für eine Stichprobe $x = (x_1, \dots, x_n)$ seien $x_{(1)}, \dots, x_{(n)}$ die der Größe nach geordneten Werte. Es gilt also

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

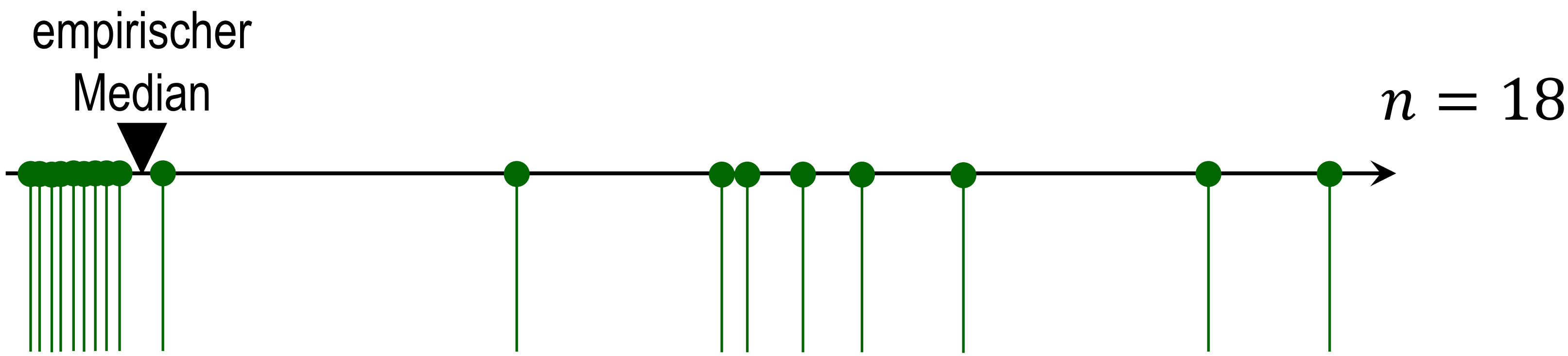
Dann heißt der durch diese geordnete Stichprobe eindeutig bestimmte Zahlenwert

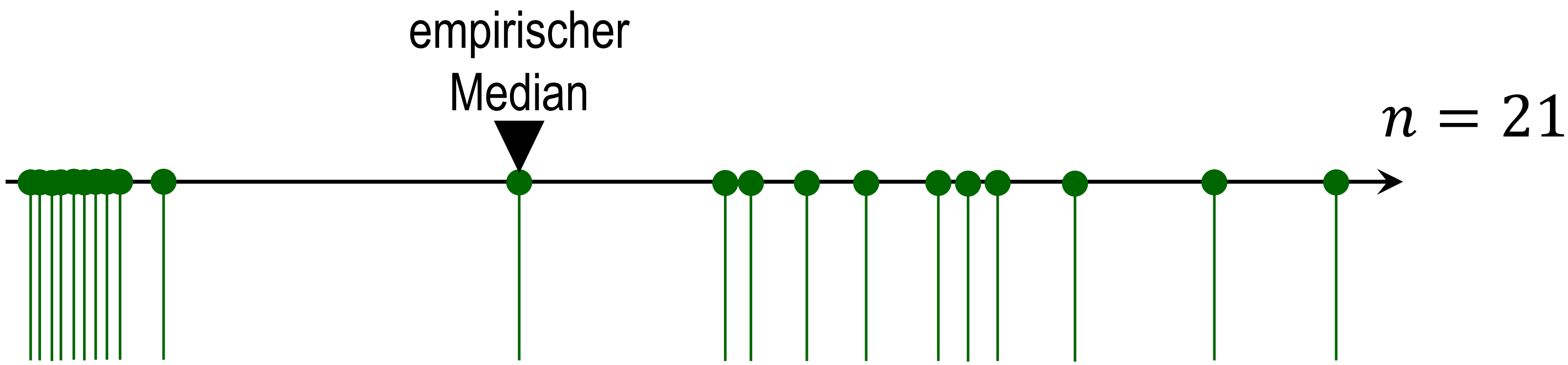
$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & falls\ n\ ungerade \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+1}{2}\right)}}{2}, & falls\ n\ gerade \end{cases}$$

der (**empirische**) **Median** (bzw. **Zentralwert**) der Stichprobe.

Eigenschaft: Der empirische Median ist sowohl größer-gleich als auch kleiner-gleich jeweils der Hälfte aller Werte der Stichprobe.

Visualisierung: Empirischer Median





Quantil (Lagemaß)

$\boldsymbol{p}$-Quantil (vgl. [2,p.65])

Sei $x = (x_1, \dots, x_n)$ eine Stichprobe. Jeder Wert x_p für den gilt

- mindestens ein Anteil p der Elemente ist kleiner-gleich x_p und zusätzlich
- mindestens ein Anteil $1 - p$ der Elemente ist größer-gleich x_p,

heißt **$\boldsymbol{p}$-Quantil** der Stichprobe.

Wichtige Beispiele:

- Der Median $\tilde{x}$ ist ein 50%-Quantil $x_{0.5}$.
- Das untere Quartil ist ein 25 %-Quantil $x_{0.25}$.
- Das obere Quartil ist ein 75 %-Quantil $x_{0.75}$.

Modalwert (Lagemaß)

Modus

Jeder Merkmalswert, der in einer Stichprobe am häufigsten vorkommt heißt **Modalwert** (bzw. **Modus**) der Stichprobe.

Beispiele:

- Der Modalwert der Stichprobe (1,3,5,7,3,5,3,6,7,8,6,4,3) ist 3.
- Für die Stichprobe (1,1,1,1,2,2,2,2,3,4,4,4,4) sind 1,2,4 Modalwerte

Bemerkung:

Der Modalwert ist nicht immer eindeutig definiert und nur für diskrete Merkmale (ohne Weiteres) sinnvoll definiert.

Geometrisches Mittel (Lagemaß)

Geometrische Mittel

Für eine Stichprobe $x = (x_1, \dots, x_n)$ heißt

$$\bar{x}_{geom} = \sqrt[n]{x_1 \cdot \ldots \cdot x_n}$$

das **geometrische Mittel** der Stichprobe.

Beispiel: Geometrisches Mittel

Eine Bakterienpopulation wächst

- am ersten Tag um den Faktor 2,
- am zweiten Tag um den Faktor 3 und
- am dritten Tag um den Faktor 4.

a) Berechnen Sie das arithmetische und das geometrische Mittel des Wachstumsfaktors.

b) Gegeben sei die eine Gesamtpopulation von 1000 Bakterien.
   - Berechnen Sie das gesamte Wachstum innerhalb dieses Zeitraums.
   - Berechnen Sie das Wachstum unter Annahme des arithmetischen Mittels als Wachstumsfaktor an den drei Tagen.
   - Berechnen Sie das Wachstum unter Annahme des geometrischen Mittels als Wachstumsfaktor an den drei Tagen.

Harmonisches Mittel (Lagemaß)

Harmonische Mittel (vgl. [6,p.35])

Sei $x = (x_1, \dots, x_n)$ eine Stichprobe, dann heißt

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

das **harmonische Mittel** der Stichprobe.

Beispiel: Arithmetisches und harmonisches Mittel

- Ein PKW fährt eine Stunde lang mit 80 km/h und dann eine Stunde mit 120km/h. Berechnen Sie die Durchschnittsgeschwindigkeit.

- Ein PKW fährt 100 km mit 80 km/h und weitere 100 km mit 120 km/h. Berechnen Sie die Durchschnittsgeschwindigkeit.

Inhaltsverzeichnis VE 03

VE 03: Lage- und Streuungsmaße

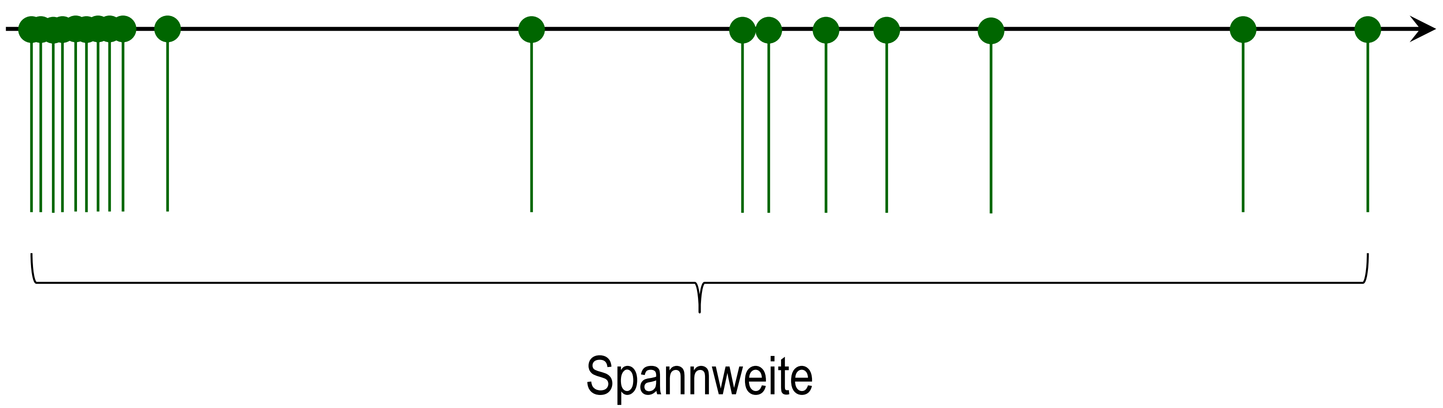
Spannweite (Streuungsmaß)

Spannweite

Sei $x = (x_1, \dots, x_n)$ eine Stichprobe. Die Differenz

$$R = \max_{i \in \{1,\dots,n\}} x_i - \min_{i \in \{1,\dots,n\}} x_i$$

heißt die **Spannweite** der Stichprobe.



Mittlere absolute Abweichung

Sei $x = (x_1, \ldots, x_n)$ eine Stichprobe vom Umfang n mit dem empirischen Mittelwert $\overline{x}$ und dem empirischen Median $\tilde{x}$.

Die **mittlere absolute Abweichung bezüglich des Mittelwerts** $\overline{x}$ ist definiert durch

$$d_{\overline{x}} = \frac{1}{n} \sum_{i=1}^{n} |x_i - \overline{x}| \, .$$

Die **mittlere absolute Abweichung bezüglich des Medians** $\tilde{x}$ ist definiert durch

$$d_{\tilde{x}} = \frac{1}{n} \sum_{i=1}^{n} |x_i - \tilde{x}| \, .$$

Beispiel: Mittlere absolute Abweichung bezüglich Mittelwert/Median

Für einen neues Automodell der gehobenen Klasse ermittelt der ADAC jeweils den genauen Benzinverbrauch auf 100 km bei unterschiedlichen Streckenprofilen:

Streckenprofil	1	2	3	4	5	6	7	8
Verbrauch in l/100 km	7.3	9.5	8.4	6.9	10.3	10.2	8.5	7.9

Empirischer Mittelwert: 8.625 l/100km

Empirischer Median: 8.45 l/100km

Mittlere absolute Abweichung bezüglich des Mittelwerts: 1.03125 l / 100 km

Mittlere absolute Abweichung bezüglich des Medians: 1.0 l / 100 km

Empirischer Varianz (Streuungsmaß)

Varianz

Sei $x = (x_1, \ldots, x_n)$ eine Stichprobe vom Umfang n mit $n \geq 2$ und dem empirischen Mittelwert $\overline{x}$, dann heißt

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

die **(empirische) Varianz** der Stichprobe x.

Berechnung mit Häufigkeiten

Kommen in der Stichprobe $x = (x_1, \ldots, x_n)$ genau die Werte $y_1, y_2, \ldots, y_k$ vor und bezeichnet $h(y_i)$ die relative Häufigkeit des Wertes y_i für $i = 1, \ldots, k$, so kann die (empirische) Varianz der Stichprobe x auch wie folgt berechnet werden:

$$s^2 = \sum_{i=1}^{k} (y_i - \overline{x})^2 \cdot h(y_i)$$

Alternative empirischer Varianz (Streuungsmaß)

Stichprobenvarianz

Sei $x = (x_1, \ldots, x_n)$ eine Stichprobe vom Umfang n mit $n \geq 2$ und dem empirischen Mittelwert $\overline{x}$, dann wird die (empirische) Varianz der Stichprobe x alternativ oftmals auch als

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

definiert. Wir nennen diese **Stichprobenvarianz**.

Empirischer Standardabweichung (Streuungsmaß)

Standardabweichung

Sei $x = (x_1, \dots, x_n)$ eine Stichprobe vom Umfang n mit $n \geq 2$ und empirischer Varianz s^2 der Stichprobe x. Die positive Quadratwurzel der empirischen Varianz

$$s = +\sqrt{s^2}$$

heißt die **(empirische) Standardabweichung** der Stichprobe x.

<u>Bemerkung</u>:

Die empirische Standardabweichung gewichtet größere Abweichungen vom Mittelwert stärker als die mittlere absolute Abweichung.

Beispiel: Empirischer Varianz und Standardabweichung

Für ein neues Automodell der gehobenen Klasse ermittelt der ADAC jeweils den genauen Benzinverbrauch auf 100 km bei unterschiedlichen Streckenprofilen:

Streckenprofil	1	2	3	4	5	6	7	8
Verbrauch in l/100 km	7.3	9.5	8.4	6.9	10.3	10.2	8.5	7.9

Empirischer Mittelwert: 8.625 l/100km

Empirischer Median: 8.45 l/100km

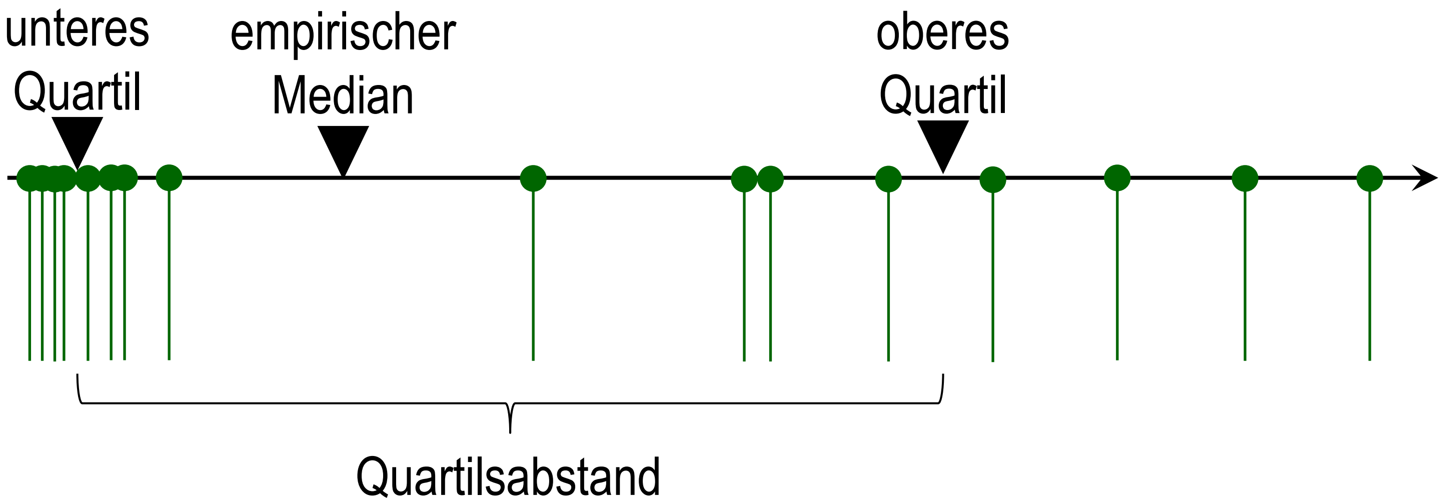
Empirische Varianz: 1.625

Empirische Standardabweichung: 1.274754878

Quartilsabstand (Streuungsmaß)

Quartilsabstand (vgl. [5,p.40])

Sei $x = (x_1, \ldots, x_n)$ eine Stichprobe vom Umfang n. Die Differenz $x_{0.75} - x_{0.25}$ zwischen dem oberen und unteren Quartil heißt **Quartilsabstand**.



<u>Bemerkung</u>:

Der Quartilsabstand beschreibt die Breite des Intervalls, in dem die mittleren 50% der Stichprobenwerte liegen.