Einfache Grundlagen der Statistik

W. Kippels

5. November 2022

Inhaltsverzeichnis

1 Vorwort

Diese und ähnliche Anleitungen zu erstellen erfordert sehr viel Zeit und Mühe. Trotzdem stelle ich alles kostenfrei der Allgemeinheit zur Verfügung. Wenn Sie diese Datei hilfreich finden, dann bitte ich Sie um Erfüllung des nachfolgend beschriebenen „Generationenvertrages“:

> *Wenn Sie später einmal Ihre Ausbildungsphase beendet haben und im Beruf stehen (oder auch noch danach), geben Sie bitte Ihr Wissen in geeigneter Form an die nachfolgende Generation weiter.*

Wenn Sie mir eine Freude machen wollen, dann schreiben Sie mir bitte eine kleine Email an die folgende Adresse: mail@dk4ek.de

Vielen Dank!

2 Einführung

Wenn irgendwo viele Daten anfallen, können Methoden der Statistik Ordnung in die Datenflut bringen und eine gewisse Übersicht schaffen. Am Beispiel der Ergebnisse beim Weitsprung im Sportunterricht einer Schulklasse sollen in diesem Artikel die Begriffe und die Rechenformeln erläutert werden. Die Ergebnisliste der Klasse sieht wie folgt aus:

Name:	**Sprungweite in Metern:**
Alina	3,87
Beate	4,12
Chantal	3,31
Dominik	3,19
Eileen	3,03
Fabian	3,80
Gisela	3,01
Henry	3,97
Isabel	4,34
Julian	3,71
Karina	3,77
Louis	4,03
Max	3,98
Nina	2,97
Oktay	3,44
Pascal	4,05
Quentin	3,25
Ralf	3,09
Simone	4,02
Tim	4,37
Ulf	3,31
Vanessa	3,75
Winfried	3,22
Xaver	3,98
Yilmaz	3,80
Zara	3,21

Wie man leicht erkennen kann, ist diese Liste nach den Namen in alfabetischer Reihenfolge sortiert. Sie kann aber durchaus auch in völlig unsortierter Reihenfolge vorliegen. Die Sortierung spielt keine Rolle.

3 Der Maximalwert

Der Maximalwert ist einfach der **größte vorkommende** Wert. Man muss nur alle vorkommenden Werte miteinander vergleichen und den größten Wert auswählen. In unserem Beispiel ist das der Sprung von Tim mit 4,37 Metern.

4 Der Minimalwert

Der Minimalwert ist der **kleinste vorkommende** Wert. Ähnlich wie beim Maximalwert muss man auch hier die Liste durchsuchen, bis man den kleinsten Wert gefunden hat. Hier ist das Nina mit ihrem Sprung mit 2,97 Metern.

5 Der Mittelwert

Der Mittelwert muss berechnet werden. Dazu addiert man alle Einzelwerte und dividiert die Summe durch die Anzahl der Einzelwerte.

Nennt man die Anzahl der Einzelwerte n und die Einzelwerte a_1, a_2, a_3, $\ldots$, a_n, dann kann man den Mittelwert durch eine Formel ausdrücken:

$$M = \frac{a_1 + a_2 + a_3 + \ldots + a_n}{n}$$

Führen wir die Rechnung hier für unser Beispiel einmal durch. Es gibt 26 Schüler und Schülerinnen in der Klasse, also ist $n = 26$. Die einzelnen Sprungweiten a_1 bis a_{26} entnehmen wir der Tabelle.

Aus Platzgründen addiere ich zunächst vorab alle einzelnen Sprungweiten:

$$\begin{aligned}
\Sigma &= a_1 + a_2 + a_3 + \ldots + a_{26} \\
&= 3{,}87 + 4{,}12 + 3{,}31 + 3{,}19 + 3{,}03 + 3{,}80 + 3{,}01 + 3{,}97 + 4{,}34 + \ldots \\
&\quad \ldots + 3{,}71 + 3{,}77 + 4{,}03 + 3{,}98 + 2{,}97 + 3{,}44 + 4{,}05 + 3{,}25 + 3{,}09 + \ldots \\
&\quad \ldots + 4{,}02 + 4{,}37 + 3{,}31 + 3{,}75 + 3{,}22 + 3{,}98 + 3{,}80 + 3{,}21 \\
\Sigma &= 94{,}95
\end{aligned}$$

Hiermit kann der Mittelwert berechnet werden:

$$M = \frac{\Sigma}{n} = \frac{94{,}95}{26} \approx 3{,}652$$

Man sagt dann: Die Schüler sind im Mittel ungefähr 3,642 Meter weit gesprungen.

6 Der Medianwert

Der Medianwert ist ein Wert, zu dem gleich viele Ergebniswerte oberhalb und unterhalb liegen. Etwas genauer formuliert: Ein Wert heißt Medianwert, wenn sowohl die Hälfte aller Werte **größer oder gleich** als auch die Hälfte aller Werte **kleiner oder gleich** diesem Wert sind.

Der Median darf **nicht** mit dem Mittelwert (siehe oben) verwechselt werden. Um den Medianwert zu ermitteln, sollte man zweckmäßigerweise alle Ergebniswerte der Größe nach sortieren. Die sortierte Liste sieht dann so aus:

Name:	**Sprungweite in Metern:**
Nina	2,97
Gisela	3,01
Eileen	3,03
Ralf	3,09
Dominik	3,19
Zara	3,21
Winfried	3,22
Quentin	3,25
Chantal	3,31
Ulf	3,31
Oktay	3,44
Julian	3,71
Vanessa	3,75
Karina	3,77
Fabian	3,80
Yilmaz	3,80
Alina	3,87
Henry	3,97
Max	3,98
Xaver	3,98
Simone	4,02
Louis	4,03
Pascal	4,05
Beate	4,12
Isabel	4,34
Tim	4,37

In unserer Beispielklasse gibt es 26 Schüler. Der Medianwert ist demnach ein Wert, den jeweils die Hälfte der Schüler (also 13) höchstens bzw. mindestens erreicht haben. Die waagerechte Trennlinie in der sortierten Liste markiert diese Stelle. Demnach wäre jede Sprungweite zwischen 3,75 und 3,77 Meter ein Medianwert. Allerdings nimmt man in einem solchen Fall in der Regel das **arithetische Mittel** der beiden benachbarten Werte als Medianwert, hier also:

$$Me = \frac{a_{13} + a_{14}}{2} = \frac{3{,}75\,\mathrm{m} + 3{,}77\,\mathrm{m}}{2} = 3{,}76\,\mathrm{m}$$

Trotzdem ist **jeder** Wert im Bereich $3{,}75\,\mathrm{m} \ldots 3{,}77\,\mathrm{m}$ ein Medianwert.

Ist die Anzahl n eine **ungerade** Zahl, dann ist der Medianwert ohne weitere Berechnung genau der mittlere Wert der sortierten Liste. Dabei kann es durchaus vorkommen, dass mehrere Werte übereinstimmen. Der Medianwert stimmt dann eventuell mit einem (oder sogar mehreren) Nachbarwerten überein. In unserem Beispiel war die Anzahl mit $n = 26$ aber eine gerade Zahl.

Vergleicht man den Mittelwert und den Medianwert in unserem Beispiel miteinander, dann stellt man fest, dass diese **nicht** übereinstimmen. In der Regel ist das immer so. Hier ist der Medianwert mit 3,73 m etwas größer, als der Mittelwert mit 3,642 m. Der Unterschied ist zwar nicht groß, es sind aber nur 12 Sprungweiten kleiner, als der Mittelwert, wogegen 14 größer sind. Julian liegt mit seiner Sprungweite über dem Mittelwert, aber unterhalb des Medianwertes.

Um den Unterschied zwischen diesen beiden Kennwerten noch etwas deutlicher zu machen, stellen wir uns zwei Schulklassen mit je 25 Schülern vor. Beide Klassen haben eine Klassenarbeit geschrieben. Dies sind die Ergebnisse der Klassenarbeiten:

Klasse A:

Note:	**Anzahl:**
1	13
2	0
3	0
4	0
5	0
6	12

Klasse B:

Note:	**Anzahl:**
1	12
2	0
3	0
4	0
5	0
6	13

In beiden Klassen kommen nur die Noten 1 und 6 vor. Weil die Anzahl der Schüler mit 25 eine ungerade Zahl ist, stellt in einer nach Notenergebnissen sortierten Liste die Note des Schülers Nummer 13 den Medianwert dar. Vor ihm und nach ihm kommen genau 12 Schüler. Demnach liegt in der Klasse A der Median bei 1 und in der Klasse B bei 6. Dieser große Unterschied mag überraschen, sieht doch auf den ersten Blick der Ausfall der Arbeit in beiden Klassen sehr ähnlich aus.

Berechnen wir nun die Mittelwerte.

$$\text{Klasse A: } M = \frac{1 \cdot 13 + 6 \cdot 12}{25} = 3{,}4$$

$$\text{Klasse B: } M = \frac{1 \cdot 12 + 6 \cdot 13}{25} = 3{,}6$$

Die Mittelwerte liegen nicht weit auseinander. Offenbar ist der Mittelwert zur Bewertung des Ausfalles einer Klassenarbeit aussagekräftiger, als der Medianwert. Das das nicht bei jedem Anwendungsfall so sein muss, zeigt das nächste Beispiel.

In der nachfolgenden Liste sind die jeweiligen Tageshöchsttemperaturen einer Woche dargestellt.

Montag	17°C
Dienstag	18°C
Mittwoch	17°C
Donnerstag	17°C
Freitag	29°C
Samstag	20°C
Sonntag	18°C

Es soll wieder der Medianwert und der Mittelwert bestimmt werden.

Damit der Medianwert besser bestimmt werden kann, werden die Messwerte nach deren Größe umsortiert.

Montag	17°C
Mittwoch	17°C
Donnerstag	17°C
Dienstag	18°C
Sonntag	18°C
Samstag	20°C
Freitag	29°C

In der Mitte steht der Dienstag mit 18°C. Diese 18°C stellen somit den Medianwert dar.

Es folgt der Mittelwert, der berechnet werden muss.

$$\vartheta_M = \frac{17°\mathrm{C} + 18°\mathrm{C} + 17°\mathrm{C} + 17°\mathrm{C} + 29°\mathrm{C} + 20°\mathrm{C} + 18°\mathrm{C}}{7} \approx 19{,}4°\mathrm{C}$$

Der „Ausreißer“ am Freitag mit 29°C zieht den Mittelwert hoch. Tatsächlich war es meistens kälter. Hier beschreibt der Medianwert von 18°C eher den Wochenverlauf der Temperatur, weil es nur an zwei Tagen wärmer war, als der Mittelwert von 19,4°C sagt.

Ein anderes Beispiel, bei dem eher der Medianwert interessant ist, ist die Lebenserwartung eines Menschen. Zunächst würde man das Alter von hinreichend vielen Menschen in einer Liste erfassen, in dem diese Menschen gestorben sind. Noch lebende Menschen kommen nicht in diese Liste. Wenn man hier den Mittelwert berechnet, dann ziehen einige wenige Menschen, die bereits in jungen Jahren gestorben sind, den Durchschnitt kräftig nach unten. Das liegt daran, dass bei diesen Menschen der Abstand zwischen dem erreichten Lebensalter und dem Durchschnittswert recht groß ist. Die Gründe für den vorzeitigen Tod sind in der Regel Unfälle, manchmal auch seltene schwere Erkrankungen. Der Medianwert ist (bekanntlich) der Wert, der nur von **genau** der Hälfte aller Menschen überschritten wird. Daher gibt dieser besser die vermutliche Lebenserwartung eines noch lebenden Menschen an.

Welche Werte man für welchen Anwendungsfall benötigt, ist allerdings eher ein philosophisches als ein mathematisches Probem. In diesem Artikel möchte ich mich ausschließlich um den mathematischen Hintergrund kümmern.

7 Übungsaufgaben

Hier folgen nun einige Beispiele. Lösungen dazu finden Sie im nächsten Kapitel.

7.1 Aufgabe 1

Bei einem erkrankten Menschen wurde regelmäßig die Körpertemperatur gemessen. Die Temperaturwerte sind in nachfolgender Liste zusammengetragen.

Zeitpunkt:	**Temperatur in ℃**
Montag morgen	37,9
Montag abend	38,2
Dienstag morgen	38,2
Dienstag abend	38,8
Mittwoch morgen	38,4
Mittwoch abend	38,5
Donnerstag morgen	38,1
Donnerstag abend	38,2
Freitag morgen	37,6
Freitag abend	37,4
Samstag morgen	36,8
Samstag abend	36,9
Sonntag morgen	36,4
Sonntag abend	36,6

Bestimmen Sie für jede Messung:

- den Maximalwert
- den Minimalwert
- den Mittelwert
- den Medianwert

7.2 Aufgabe 2

Das Ergebnis der Klassenarbeit einer fiktiven Klasse sieht wie folgt aus:

Name:	**erreichte Punkte:**	**Note:**
Alessio	89	1
Carina	61	3
Elisa	78	2
Fabian	47	4
Jessica	35	5
Julia	69	3
Juliana	39	4
Kristina	55	3
Lukas	43	4
Maximilian	35	5
Karina	47	4
Matthias	95	1
Melina	59	3
Sabrina	93	1
Sahra	68	3

1. Bestimmen Sie die beste Note! Wer hat die?
2. Bestimmen Sie den Maximalwert der Punktezahlen. Wer hat die?
3. Bestimmen Sie die schlechteste Note! Wer hat die?
4. Bestimmen Sie den Minimalwert der Punktezahlen. Wer hat die?
5. Bestimmen die die Durchschnittsnote.
6. Bestimmen Sie die durchschnittlich erreichte Punktezahl.
7. Bestimmen Sie einen Medainwert für die Noten.
8. Bestimmen Sie einen Medianwert für die Punktezahlen.

8 Lösungen

8.1 Aufgabe 1

Der Maximalwert von Dienstag abend beträgt: $\vartheta_{max} = 38{,}8\,°\mathrm{C}$
Der Minimalwert von Sonntag morgen beträgt: $\vartheta_{min} = 36{,}4\,°\mathrm{C}$

Der Mittelwert wird berechnet:

$$\begin{aligned}
\vartheta_M &= \frac{\vartheta_1 + \vartheta_2 + \ldots + \vartheta_{14}}{14} \\
&= \frac{37{,}9 + 38{,}2 + 38{,}2 + 38{,}8 + 38{,}4 + 38{,}5 + 38{,}1 + 38{,}2 + 37{,}6 + 37{,}4 + 36{,}8 + 36{,}9 + 36{,}4 + 36{,}6}{14} \\
&= \frac{538}{14} \\
\vartheta_M &\approx 38{,}43
\end{aligned}$$

Die Durchschnittstemperatur betrug ungefähr $\vartheta_M \approx 38{,}43°\mathrm{C}$.

Zur Bestimmung des Medianwertes sortiere ich die Tabelle nach der gemessenen Temperatur.

Zeitpunkt:	**Temperatur in °C**
Sonntag morgen	36,4
Sonntag abend	36,6
Samstag morgen	36,8
Samstag abend	36,9
Freitag abend	37,4
Freitag morgen	37,6
Montag morgen	37,9
Donnerstag morgen	38,1
Donnerstag abend	38,2
Montag abend	38,2
Dienstag morgen	38,2
Mittwoch morgen	38,4
Mittwoch abend	38,5
Dienstag abend	38,8

8.2 Aufgabe 2

zu 1: Alessio, Matthias und Sabrina haben eine 1 als beste Note.

zu 2: Matthias hat mit 95 Punkten die höchste Punktezahl.

zu 3: Jessica und Maximilian haben mit einer 5 die schlechteste Note.

zu 4: Jessica und Maximilian haben mit 35 Punkten die niedrigste Punktezahl.

zu 5: Dazu berechnen wir den Mittelwert M_N der Noten mit der zugehörigen Formel. Da in der Klasse 15 Schülerinnen und Schüler sind, ist $n = 15$.

$$\begin{aligned} M_N &= \frac{N_1 + N_2 + \ldots + N_{15}}{15} \\ &= \frac{1+3+2+4+5+3+4+3+4+5+4+1+3+1+3}{15} \\ &= \frac{43}{15} \\ M_N &\approx 2{,}87 \end{aligned}$$

Die Durchschnittsnote ist ungefähr 2,87.

zu 6: Nach dem selben System berechnen wir den Mittelwert M_P für die erreichten Punktezahlen.

$$\begin{aligned} M_P &= \frac{P_1 + P_2 + \ldots + P_{15}}{15} \\ &= \frac{89+61+78+47+35+69+39+55+43+35+47+95+59+93+68}{15} \\ &= \frac{913}{15} \\ M_P &\approx 60{,}87 \end{aligned}$$

Die durchschnittlich erreichte Punktezahl beträgt ungefähr 60,87.

zu 7 und 8: Um die Medianwerte besser bestimmen zu können sortiere ich die Liste nach den Punktezahlen. Dabei ist sie dann automatisch auch nach Noten sortiert.

Name:	**erreichte Punkte:**	**Note:**
Matthias	95	1
Sabrina	93	1
Alessio	89	1
Elisa	78	2
Julia	69	3
Sahra	68	3
Carina	61	3
Melina	59	3
Kristina	55	3
Fabian	47	4
Karina	47	4
Lukas	43	4
Juliana	39	4
Jessica	35	5
Maximilian	35	5

Es sind 15 Schülerinnen und Schüler in der Klasse. Das ist eine **ungerade** Zahl. In der Mitte (zwischen den beiden horizontalen Linien) steht Melina mit 59 Punkten und der Note 3. Der Medianwert für die Punktezahlen ist daher 59 und der Medianwert für die Noten ist die 3.

Wir können das auch rechnerisch nachprüfen. Die Hälfte von 15 (Gesamtschülerzahl) ist 7,5. Beginnen wir mit der Punktezahl. Nach der Definition des Medianwertes müssen mindestens 7,5 erreichte Punktezahlen **mindestens 59** betragen und auch mindestens 7,5 mal **maximal 59** sein. Prüfen wir das nach.

Es gibt 8 Einträge von 35 bis 59 Punkte, das ist tasächlich mindestens 7,5 mal. Von 59 bis 95 gibt es ebenfalls 8 Einträge. Damit ist auch hier die Mindestzahl von 7,5 erfüllt. 59 Punkte ist der Medianwert für die Punkte.

Prüfen wir nun, ob die Note 3 der Medianwert ist. Es gibt insgesamt 9 mal die Note 3 oder besser und 11 mal die Note 3 und schlechter. Damit sind die geforderten 7,5 mal für beide Gruppen erfüllt. Die 3 ist Medianwert für die Noten.