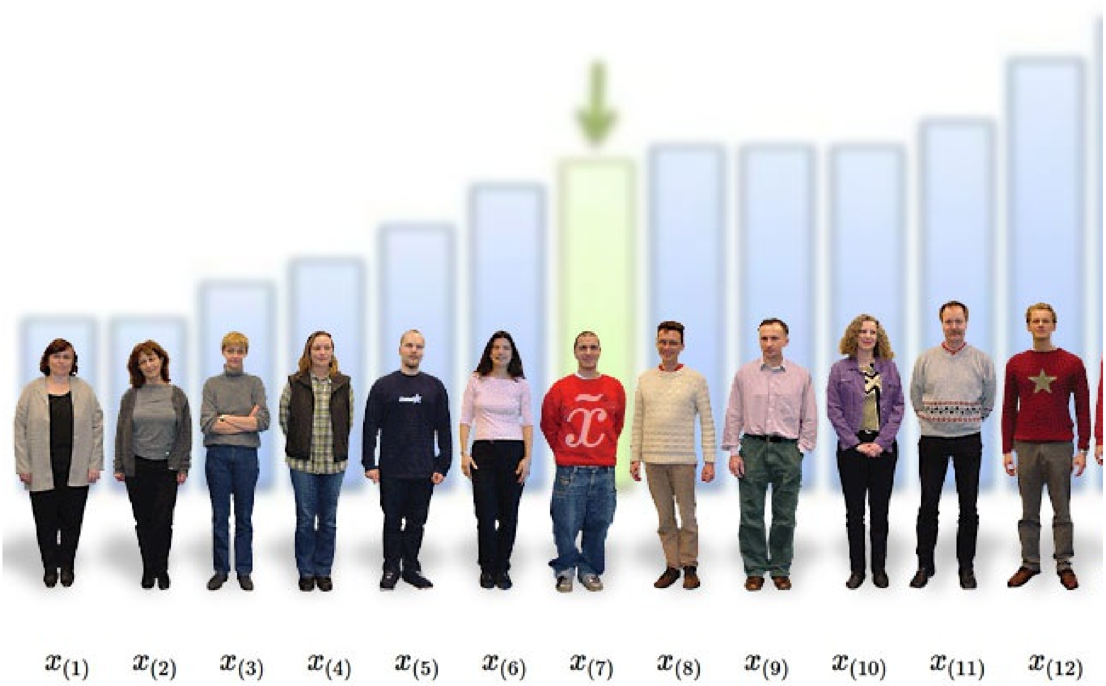
MMO - Median und Modus



Lernziele und Überblick

In dieser Lerneinheit betrachten wir weiter die Lagemaße.Sie lernen den Median zu berechnen und erfahren mehr über seine Eigenschaften. Zusätzlich werden Sie in dieser Lerneinheit ein weiteres Lagemaß, den Modus kennenlernen.

Lernziele

Nach dem Durcharbeiten dieser Lerneinheit sollen Sie

- die Begriffe Median und Modus erklären können,
- den Median und Modus berechnen können,
- die Unterschiede und Gemeinsamkeiten von Median und arithmetischem Mittel erklären können
- die Minimum- und die Transformationseigenschaft des Medians beschreiben und anwenden können,

Gliederung der Lerneinheit

Zeitbedarf und Umfang

Für die Durcharbeitung dieser Lerneinheit benötigen Sie ca. 90 Minuten und für die Übungen mit der Statistiksoftware **R** ca. 120 Minuten.

1 Der Median - in Mitten der Daten

Den Median haben sie bereits im Rahmen der Lerneinheit *„QBX - Quantile und Boxplots“* kennengelernt. Der Median ist das 0,5-Quantil. Er teilt einen geordneten Datensatz in zwei Hälften mit gleichem Datenumfang. Zwei Eigenschaften zeichnen den Median besonders aus, zum einen ist er im Gegensatz zum geometrischen und arithmetischen Mittel auch für nur ordinal skalierte Variablen geeignet zum anderen ist er robust gegen Ausreißer. Schauen wir ihn uns mal genauer an.

Wenn unsere Daten mindestens ordinal skaliert sind, lassen sie sich der Größe nach ordnen.

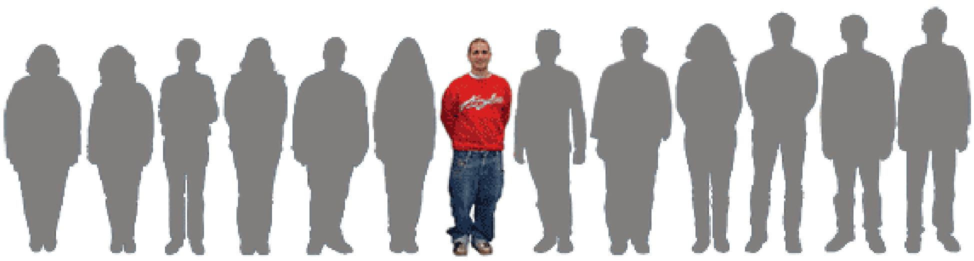
$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq x_{(4)} \leq x_{(5)} \leq x_{(6)} \leq x_{(7)} \leq x_{(8)} \leq x_{(9)} \leq x_{(10)} \leq x_{(11)} \leq x_{(12)} \leq x_{(13)}$$

Abb.: Sortierung nach Größe

Aus den geordneten Werten kann man leicht Informationen über die Verteilung und die Lage der Daten sammeln. Dabei ist häufig der mittlere Wert interessant. Wir suchen also den Wert, der die nach der Größe geordneten Merkmalausprägungen halbiert, so dass gleich viele Messwerte links und rechts des Wertes liegen. Dieser Wert wird als Median bezeichnet.

Das Formelzeichen für den Median ist $\tilde{x}_{0,5}$ oder schlicht $\tilde{x}$ also ein x mit einer Tilde darüber. Bei der Ermittlung des Medians eines Datensatzes muss zwischen gerader und ungerader Anzahl von Werten unterschieden werden.

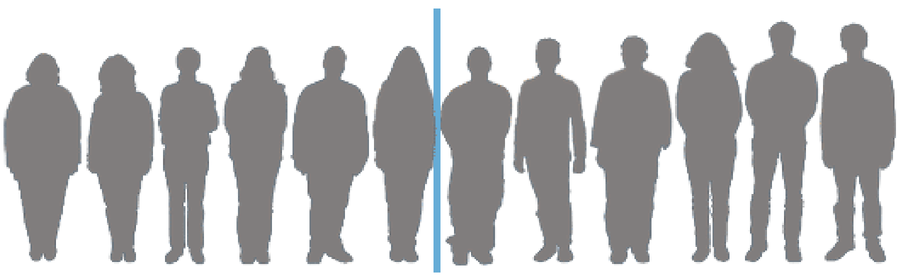
Bei einer **ungeraden** Anzahl n ist der Median $x_{\left(\frac{n+1}{2}\right)}$ genau der Wert in der Mitte. Er ist also einer der ursprünglichen Datenwerte.



$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq x_{(4)} \leq x_{(5)} \leq x_{(6)} \leq x_{(7)} \leq x_{(8)} \leq x_{(9)} \leq x_{(10)} \leq x_{(11)} \leq x_{(12)} \leq x_{(13)}$$

Abb.: Median bein n=13 (ungerade)

Bei **geradem** n wird die Mitte zwischen $x_{\left(\frac{n}{2}\right)}$ und $x_{\left(\frac{n}{2}+1\right)}$ zum Median erklärt. Der Median wird also interpoliert. Hier ist der Median keiner unserer ursprünglichen Werte sondern liegt dazwischen.



$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq x_{(4)} \leq x_{(5)} \leq x_{(6)} \leq x_{(7)} \leq x_{(8)} \leq x_{(9)} \leq x_{(10)} \leq x_{(11)} \leq x_{(12)}$$

Abb.: Median bei n=12 (gerade)

1.1 Der Median (Definition)

Definition

Median

Ist X ein mindestens ordinales skaliertes Merkmal, mit den nach Größe geordneten Merkmalsausprägungen $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ dann ist der Median definiert durch:

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

Die Formel stimmt natürlich mit der zur allgemeinen Berechnung von Quantilen überein. Diese kennen Sie bereits aus der Lerneinheit QBX Kapitel 2

Betrachten wir zwei einfache Rechenbeispiele.

Beispiel

Gegeben ist die **erste** Stichprobe mit 3, 7, 5, 9, 1, 0, 4, 11, 1

1. Ordnen der Werte: 0, 1, 1, 3, 4, 5, 7, 9, 11
2. Umfang n=9 - also **ungerade**!
3. Median liegt an der Stelle = $\frac{9+1}{2} = 5$
4. Median $\tilde{x} = x_{(5)} = 4$

Beispiel

Gegeben ist die **zweite** Stichprobe mit 12, 18, 18, 11, 10, 8, 15, 20, 19, 13

1. Ordnen der Werte: 8, 10, 11, 12, 13, 15, 18, 18, 19, 20
2. Umfang n=10 - also **gerade**!
3. Median = $\frac{x_{(5)}+x_{(6)}}{2}$
4. Median $\tilde{x} = \frac{13+15}{2} = 14$

Je nachdem, ob wir es mit einer geraden oder ungeraden Anzahl von Daten zu tun haben, steht ein Wert entweder genau in der Mitte, oder der Median wird aus beiden mittleren Werten gemittelt. Die Berechnung bei geradem n gilt streng genommen nur für metrische Daten, denn der Durchschnitt der beiden mittleren Werte ist für ordinale Daten nicht berechenbar. Hier helfen wir uns pragmatisch wie folgt: Sind die beiden mittleren Werte identisch, kann der Median als dieser Wert angegeben werden; andernfalls könnte man sagen: „Der Median liegt zwischen ... und ...“

Damit Sie diese Unterscheidung im Schlaf beherrschen und ein Gefühl für den Median gewinnen, folgt nun eine Reihe von einfachen Übungen.

1.2 Reichlich Rechenübungen

Versuchen Sie diese einfachen Übungen selbst zu lösen, bevor Sie sich die Lösungen anschauen.

Berechnen

Übung MMO-01

Median bestimmen (1)

5,7	7,1	6	5,1	3,7	7,7	2,5	7,5	4,1	3,5	4,5

11

Bestimmen Sie den Median aus den gegebenen Werten.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-02

Median bestimmen (2)

Die Punktzahlen, die eine Studentin bei acht Klausuren erreichte, waren 87, 78, 97, 67, 91, 92, 74 und 79.

8

Bestimmen Sie den Median der Punktzahlen.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-03

Median bestimmen (3)

Die Monatsgehälter von acht Angestellten betrugen
1500 €, 2400 €, 1800 €, 2000 €, 2100 €, 5500 €, 2200 €, 1900 €.

8

Bestimmen Sie den Median ihrer Gehälter.

Das arithmetische Mittel für diese Daten, das wir in Übung ARM-03 (Lerneinheit ARM, Kap. 2.2) berechnet haben lag bei 2425 €. Wie erklären Sie sich den großen Unterschied zum Median?

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-04

Median bestimmen (4)

Die Monatsgehälter der Angestellten für drei Abteilungen der Firma „Lepaute &Töchter“ sind:

15

Abteilung L: 1600 €, 1850 €, 1550 €, 2200 €, 2100 €.
Abteilung M: 2500 €, 1950 €, 2250 €, 1700 €, 2300 €, 1500 €, 2100 €.
Abteilung N: 2400 €, 2350 €, 1875 €.

Bestimmen Sie den Median der Gehälter für jede Abteilung und dann für alle Abteilungen.

Was ist hier anders als beim arithmetischen Mittel?

Lösung (Siehe Anhang)

Bearbeitungszeit: 8 Minuten

Berechnen

Übung MMO-05

Median bestimmen (5)

Bestimmen Sie den Median der Werte dieses kardinalskalierten Merkmals $x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}, x_{(5)}$ und $x_{(6)}$

Gefragt ist hier eine Formel, da wir keine Zahlenwerte für $x_{(1)}, x_{(2)} \ldots$ kennen.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

1.3 Der Median bei klassierten Daten

Bei klassierten Daten muss man sich Gedanken machen, wie man die Mitte der Daten näherungsweise bestimmen kann. Der Median ist das 0.5- Quantil und wie man Quantile grundsätzlich für klassierte Daten berechnet, wurde bereits in Lerneinheit QBX, Kapitel 2.6 erklärt. Das Vorgehen und die Herleitung der Formel speziell für den Median wollen wir hier verdeutlichen.

Beispiel

Einzimmerwohnungen

Wir wollen den Median $\tilde{x}$ der Mieten in € von 99 freien Einzimmerwohnungen in Berlin auf der Grundlage der in der folgenden Tabelle vorgestellte Häufigkeitsverteilung bestimmen.

Index k	Klasse K_k	absolute Klassenhäufigkeit n_k	kumulierte absolute Klassenhäufigkeit N_k	relative Klassenhäufigkeit h_k	Verteilungsfunktion F_n bzw. kumulierte relative Häufigkeiten
1	K_1 = (0; 200]	0	0	0,00	0
2	K_1 = (200; 400]	43	43	0,43	0,43
3	K_1 = (400; 600]	38	81	0,38	0,81
4	K_1 = (600; 800]	9	90	0,09	0,90
5	K_1 = (800; 1000]	3	93	0,03	0,93
6	K_1 = (1000; 1200]	4	97	0,04	0,97
7	K_1 = (1200; 1400]	1	98	0,01	0,98
8	K_1 = (1400; 1600]	1	99	0,01	1

Tab.: Häufigkeitsverteilung von Wohnungen

Medianklasse

Bei der Berechnung des Medians für klassierten Daten ist es nicht von Interesse ob der Datenumfang gerade oder ungerade ist.

Wir betrachten die Hälfte des Stichprobenumfangs, mit n=99 also 49,5. Damit ist klar, dass der Median in der dritten Klasse, K_3, liegt, hier grün hervorgehoben. Diese Klasse nennt man daher **Medianklasse**. Das bedeutet in unserem Beispiel liegt der Median zwischen 400 € und 600 € Miete. Aber wo genau in der Klasse liegt der Median?

Wir unterstellen nun, dass sich die Mieten gleichmäßig über den Bereich der jeweiligen Klasse erstrecken. Das muss keineswegs der Fall sein und wird es auch in der Regel nicht sein, aber es ist ein Weg den Median näherungsweise zu berechnen.

Die Medianklasse hat eine Breite von 200 € und es liegen 38 Wohnungen in dieser Klasse. Bei einer gleichmäßigen Verteilung hieße das, dass die Mietpreise im Abstand von $\frac{200}{38} \approx 5{,}3\,€$ im Intervall **(400; 600]** liegen, also bei 400; 405,3; 410,6; 415,9 … .

Wir suchen die Miete an der 49,5ten Stelle. In den Klassen vor der Medianklasse liegen bereits 43 Werte, d. h. wir suchen den 6.5ten Wert in der Medianklasse. Nach unserer Berechnung zuvor ist das der Wert $\tilde{x} = 400\,€ + 6{,}5 \cdot 5{,}3\,€ \approx 434\,€$. Der tatsächliche Median der unklassierten Daten ist 423 €. Der Fehler der Näherung hält sich also in Grenzen.

Allgemein können wir die Median gruppierter bzw. klassierter Daten folgendermaßen näherungsweise berechnen:

Definition

Median bei klassierten Daten (gegeben absolute Häufigkeiten)

Der Median $\tilde{x}$ wird durch Interpolation bestimmt, so dass näherungsweise die Bedingung $F_n(\tilde{x}) = 0{,}5$ erfüllt ist. Dazu wird zunächst die Klasse ermittelt, in die der Median fällt (Medianklasse K_m) . Der Median $\tilde{x}$ wird dann wie folgt genähert:

$$\tilde{x} = u_m + \frac{\left(\frac{n}{2}\right) - N_{m-1}}{n_m} \cdot (o_m - u_m)$$

u_m oder o_m : Unter- bzw. Obergrenze der Medianklasse,
N_{m-1}: kumulierte absolute Klassenhäufigkeit der Klassen unterhalb der Medianklasse,
n_m: absolute Klassenhäufigkeit der Medianklasse
n: Stichprobenumfang.

Analog können wir auch die relativen Häufigkeiten zur Berechnung nutzen. Im Fall der Einzimmerwohnungen rechnen wir dann:

$$\tilde{x} = 400\,€ + \frac{0{,}5 - 0{,}43}{0.38} \cdot (600\,€ - 400\,€) \approx 434\,€$$

Allgemein gilt:

Definition

Median bei klassierten Daten (gegeben relative Häufigkeiten)

Der Median klassierter Daten lässt sich näherungsweise berechnen durch:

$$\tilde{x} = u_m + \frac{0{,}5 - F_{m-1}}{h_m} \cdot (o_m - u_m)$$

F_{m-1}: kumulierte relative Häufigkeit der Klassen unterhalb der Medianklasse,
h_m: relative Häufigkeit der Medianklasse

1.4 Grafische Bestimmung des Medians klassierter Daten

Neben der Berechnung des Median gibt es noch zwei grafische Möglichkeiten den Median zu bestimmen - die Verteilungsfunktion und das Histogramm.

Verteilungsfunktion

Es ist möglich den Median anhand der Grafik der Verteilungsfunktion zu interpolieren. Dazu zeichnen wir die Verteilungsfunktion der klassierten Daten und bestimmen anhand des Graphen das 0,5-Quantil und damit den Median. Wir nehmen dazu nochmals die Daten der Einzimmerwohnungen.

Die folgende Grafik zeigt das Ergebnis:

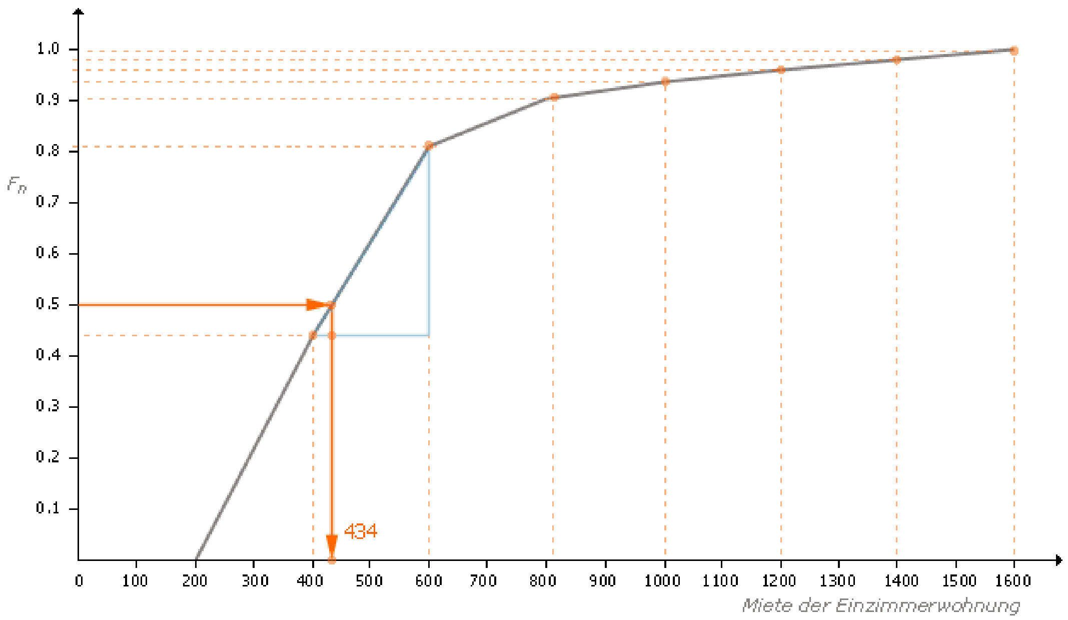


Abb.: Bestimmung des Medians (Verteilungsfunktion)

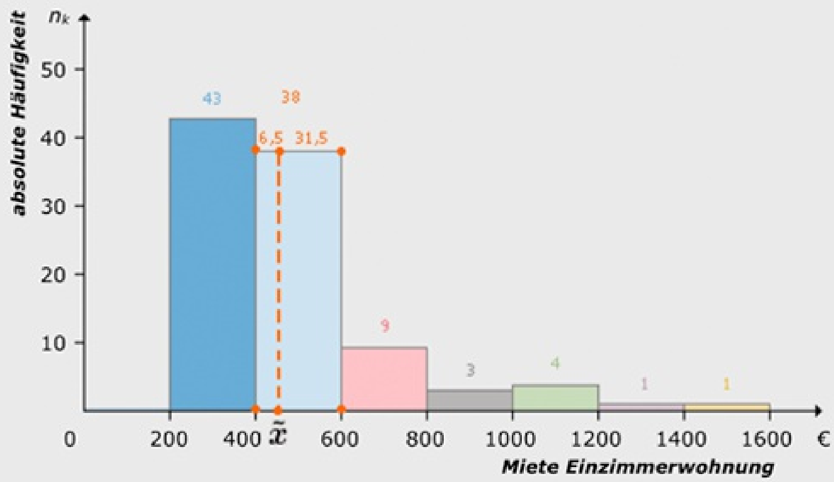
Wir lesen den Median am Graph der empirischen Verteilungsfunktion ab. Das Ergebnis ist 434 Euro - wie bei den Berechnungen zuvor.

Bestimmung aus dem Histogramm

Der Median teilt das Histogramm in zwei Teile mit gleich großem Flächeninhalt.

Rolloverbild

Bestimmung des Medians aus dem Histogramm



Die Hälfte der Fläche optisch zu erkennen ist schwierig. Man müsste nachrechen und messen. In der der Regel berechnet so niemand den Median, aber Sie sollten wissen, dass der Median die Histogramfläche genau **zwei gleichgroße Flächen** teilt.

2 Eigenschaften des Medians

Der Median ist ein Mittelwert und Lageparameter der Ihnen in der Statistik ständig begegnen wird. Sie sollten sich daher seiner Aussagekraft und Grenzen, also seiner Eigenschaften, bewusst sein – gerade wenn Sie ihn selber benutzen.

Wie wir gesehen haben, teilt der Median die Daten in genau zwei gleiche Teile. Der Median ist der Wert in der Mitte.

Der Median eignet sich für Daten, die mindestens ordinalskaliert sind.

Robustheit

Robust gegenüber extremen Werten

Der Median eignen sich für Daten, die mindestens ordinalskaliert sind. Im Gegensatz zum arithmetischen Mittel ist der Median völlig unempfindlich gegenüber extremen Werten. Alle Werte unterhalb und oberhalb des Medians können beliebig verändert werden, solange sie in „ihrer Hälfte" bleiben, ohne dass sich der Wert des Medians ändert. Daher wird der Median als **robust** bezeichnet.

Minimumeigenschaft

Der Median minimiert die Summe der absoluten Abweichungen. Die Summe der absoluten Abweichungen des Medians von den Beobachtungen $x_i, i = 1, \ldots, n$ ist minimal.

$$\sum_{i=1}^{n} |x_i - \tilde{x}_{0,5}| = \text{minimal}$$

Anders formuliert bedeutet das:

$$\boxed{\sum_{i=1}^{n} |x_i - \tilde{x}_{0,5}| \leq \sum_{i=1}^{n} |x_i - c| \text{ für alle c}}$$

Es existiert also keine Zahl, für die die Summe der absoluten Abweichungen zu den Datenwerten kleiner ist als die summierten absoluten Abweichungen der Werte zum Median.

Transformationseigenschaft

Die Bestimmung des Medians $\tilde{x}_{0,5}$ ist mit allen streng monotonen Transformationen vertauschbar.

Wir erhalten also dasselbe Ergebnis, unabhängig davon, ob wir aus den Daten zuerst den Median bestimmen und dann monoton transformieren oder ob wir die Daten erst transformieren und dann den Median bestimmen.

Monotone Transformationen schließen natürlich die linearen ein (siehe Lerneinheit *„GST - Grundbegriffe der Statistik"* im Kapitel 8.1 Arten von Transformationen).

Betrachten wir dazu zwei Beispiele.

Beispiel

Lineare Transformation

Werden die Daten $x_1, \ldots, x_n$ linear transformiert mit $y_i = a + bx_i, i = 1, \ldots, n$ dann ist der Median der transformierten Daten gleich dem transformierten Median: $\tilde{y} = a + b\tilde{x}$

Beispiel: Berechnung des Medians der transformierten Daten $y_i, i = 1, \ldots, 5$

1. x-Werte: $1, 2, 3, 4, 5$

   Median $\tilde{x} = 3$
2. $a = 4,\ b = 3,\ \ \Rightarrow y_i = 4 + 3 \cdot x_i,\ i = 1, \ldots, 5,$
3. y-Werte:
   $y_1 = 4 + 3 \cdot 1 = 7$
   $y_2 = 4 + 3 \cdot 2 = 10$
   $y_3 = 4 + 3 \cdot 3 = 13$
   $y_4 = 4 + 3 \cdot 4 = 16$
   $y_5 = 4 + 3 \cdot 5 = 19$

   Median $\tilde{y} = 13$
4. Statt die y-Werte auszurechnen, hätten Sie aber auch einfach so rechnen können:
   $\tilde{y} = a + b\tilde{x} = 4 + 3 \cdot = 13!$

Sie sehen, die Verträglichkeit des Medians mit streng monotonen Transformationen zu kennen, kann unter anderem viel Zeit beim Rechnen sparen ;-). Vergleichen Sie diese Berechnung gerne mit Lerneinheit *„ARM, Kapitel 3.3“*

Beispiel

Logarithmische Transformation (Bedingung $x \geq 0$)

$y_i = \log x_i \ \Rightarrow \ \tilde{y} = \log \tilde{x}$

x-Werte: $10, 100, 1000, 10000, 100000$

Median: $\tilde{x} = 1000$

y-Werte: $1, 2, 3, 4, 5$

Median: $\tilde{y} = 3$

Auch hier hätten wir vereinfacht rechnen können: $\tilde{y} = \log \tilde{x} = \log 1000 = 3$

3 Vergleich arithmetisches Mittel und Median

Median und arithmetisches Mittel sind beides Lagemaße mit denen auf unterschiedliche Art und Weise eine Zentrum in der Verteilung bestimmt wird. Sie sind von grundlegender Bedeutung in der Statistik. In der Praxis empfiehlt sich stets, wenn möglich, beide Parameter zu berechnen. Um das Ergebnis der eigenen Berechnungen einordnen und beurteilen zu können, ist es wichtig die Eigenschaften beider Lagemaße, auch im Vergleich, zu kennen.

1. Skalen

Der Median ist bereits für ordinalskalierte Daten geeignet während die Daten für das arithmetische Mittel mindestens intervallskaliert sein müssen.
(Siehe auch „*GST Kap. 5.5*“)

2. Minimumeigenschaften

Während das arithmetische Mittel die Summe der quadrierten Abweichungen minimiert,

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

gilt dies für den Median bezüglich der Summe der absoluten Abweichungen, d. h.

$$\sum_{i=1}^{n}|x_i - \tilde{x}_{0,5}| = \text{minimal}$$

3. Berechnung

In die Berechnung des Medians gehen im Grunde nur ein oder zwei Werte direkt ein während beim arithmetischen Mittel alle Datenwerte mit in die Berechnung einbezogen werden.

4. Robustheit

Der entscheidendste Unterschied liegt wohl in der Robustheit. Während das arithmetische Mittel sehr empfindlich auf extreme Werte, Messfehler und andere Ausreißer reagiert, bleibt der Median davon unbeeinflusst.

Das möchten wir hier anhand von zwei Beispielen verdeutlichen.

Beispiel

Erspartes

Elf Kommilitonen und Kommilitoninnen aus dem Studiengang Wirtschaftsingenieurwesen vergleichen ihre Rücklagen. Eine Studentin hat geerbt und besitzt 500.000 Euro. Die anderen 10 besitzen 0 €, 100 €, 200 €, 300 €, 400 €, 500 €, 600 €, 700 €, 800 € und 900 €.

Das arithmetische Mittel liegt 45.864 € obwohl 10 der 11 sehr viel weniger auf dem Sparbuch haben!

Der Median liegt bei 500 €, 5 Studierende haben weniger und 5 Studierende haben mehr.

Beispiel

Umsätze in der Wintersaison

Das Reiseunternehmen Hagen Alpin Tours aus dem Allgäu hat in den letzten 8 Jahren in der Wintersaison folgende Umsätze gemacht:

Jahr	Umsatz
2014	4.374.629,18 €
2015	3.758.411,59 €
2016	4.192.117,88 €
2017	4.740.855,66 €
2018	3.699.684,18 €
2019	2.867.372,37 €
2020	471.642,07 €
2021	2.997.647,75 €
2022	3.238.218,45 €

Quelle: Hagen Alpin Tours, https://www.pulver-schnee.de
© Hagen Alpin Tours, Image by gsibergerin from Pixabay

Mit Hilfe von R können wir aus den Daten schnell den Median und das arithmetische Mittel berechnen. Die folgende Abbildung stellt das Ergebnis grafisch dar.

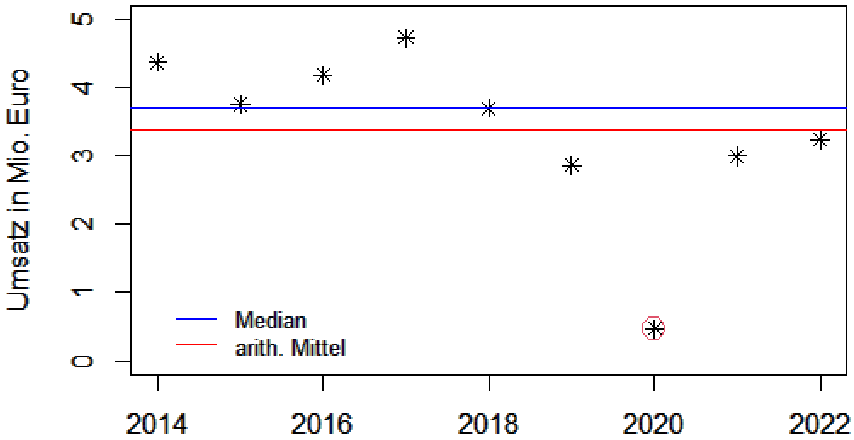


Abb.: Umsätze Hagen Alpintours Wintersaison

Wie Sie sehen, liegt das arithmetische Mittel unterhalb des Medians. Während das arithmetische Mittel nur 3.371.175 € beträgt, erreicht der Median 3.699.684 €, ein Unterschied von 328.509 €. Der Grund ist natürlich das „Corona-Jahr" 2020. Viele Unternehmen hatten in dieser Zeit mit wirtschaftlichen Einbußen zu kämpfen. Durch die zeitweise sehr strengen Reisebeschränkungen und weiterer Auflagen war besonders die Reisebranche betroffen. Für das Unternehmen Hagen Alpin Tours sieht man die Folgen in der Grafik sehr deutlich an dem ungewöhnlich geringen Umsatz im Winter 2020. Der Median bleibt weitgehend unbeeinflusst von solchen **Ausreißern**, während das arithmetische Mittel empfindlich reagiert, dies wird deutlich wenn der Wert des Jahres 2020 bei der Berechnung nicht berücksichtigt wird:

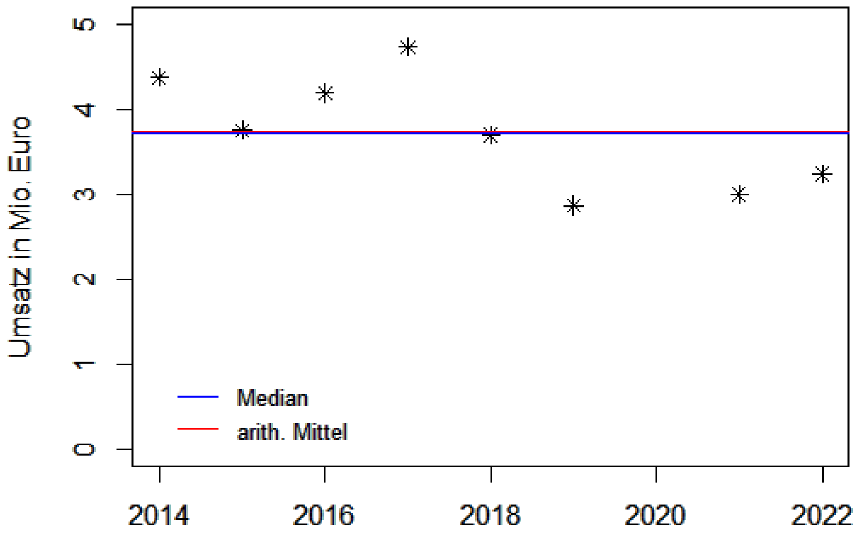


Abb.: Umsätze Hagen Alpintours Wintersaison ohne 2020

Entfernen wir die durch die Pandemie dezimierten Umsätze des Jahres 2020 sind Median und arithmetisches Mittel nahezu identisch. Während der Median sich nur leicht verändert, stieg das arithmetische Mittel um 362.442 Euro auf das Niveau des Medians. Hier die berechneten Ergebnisse.

	Wintersaison 2014-2022	**Wintersaison 2014-2022 ohne 2020**	**Differenz**
Arithmetisches Mittel	3.371.175 €	3.733.617 €	362.442 €
Median	3.699.684 €	3.729.048 €	29.364 €
Differenz	328.509 €	4.569 €	

Die Daten des Beispiels können Sie sich hier herunterladen.

Wintersaison.R

Wintersaison.csv (2 KB)

4 Der Modus (Modalwert)

Der Modus auch der Modalwert genannt, ist der Wert, der am häufigsten in den Daten vorkommt. Er ist im Grunde der einzige, sehr einfach zu bestimmende, Lageparameter, der sich bereits für nominalskalierten Daten eignet.

Definition

Modus

Der Modus ist der in den Daten am häufigsten vorkommende Wert. Gibt es keine gehäuft auftretenden Werte, wird der Modus als nicht existent bezeichnet.

Synonyme

Synonyme für Modus sind: Modalwert, dichtester Wert, häufigster Wert.

Der Modus ist vergleichsweiße ungebräuchlich mit Ausnahme bei nominal skalierten Daten. Nichtsdestotrotz gibt es eine Vielzahl von Formelzeichen mit denen der Modus bezeichnet wird, unter anderem: $\bar{x}_M$, x_{Mod} , $\breve{x}$, h, D, etc.

Beispiel

Wohnungssuche

Im Beispiel Wohnungssuche sind ca. 40 % aller angebotenen Wohnungen in Berlin Zweizimmerwohnungen und damit die Mehrheit. Der Modus der Zimmeranzahl liegt also bei zwei Zimmern. Wohnungen mit anderen Zimmerzahlen sind dagegen vergleichsweise weniger häufig. Der häufigste Wert ist hier also typisch für die gesamte Verteilung.

Wohnungen in Berlin mit Angabe der Zimmeranzahl

Index i	Zimmer-zahl x_i	absolute Häufigkeit n_i	absolute Summenhäufigkeit N_i	relative Häufigkeit h_i	relative Summenhäufigkeit H_i
1	1	28	**28**	0,110	**0.110**
2	2	102	28 + 102 = **130**	0,402	0,110 + 0,402 = **0,512**
3	3	61	130 + 61 = **191**	0,240	0,512 + 0,240 = **0,752**
4	4	39	191 + 39 = **230**	0,154	0,752 + 0,154 = **0,906**
5	5	11	230 + 11 = **241**	0,043	0,906 + 0,043 = **0,949**
6	6	9	241 + 9 = **250**	0,035	0,949 + 0,035 = **0,984**
7	7	3	250 + 3 = **253**	0,012	0,984 + 0,012 = **0,996**
8	8	1	253 + 1 = **254**	0,004	0,996 + 0,004 = **1**

Der Modus muss aber nicht existieren und selbst wenn, so braucht er nicht eindeutig zu sein. Es gibt auch bi- oder multimodale Verteilungen. D. h. Verteilungen bei denen zwei oder mehr Werte gehäuft auftreten.

Beispiel

Modalwert: Einen Modus

Die Datenreihe 1, 2, 3, 5, 7, 7, 9, 10, 15 hat den Modus 7.

Beispiel

Modalwert: Zwei Modi

Die Datenreihe 2, 3, 3, 3, 5, 7, 7, 7, 8, 11, 15, 19 hat zwei Modi, 3 und 7, und wird als bimodal bezeichnet.

Beispiel

Modalwert: Keinen Modus

Die Datenreihe 2, 3, 4, 5, 10, 12, 17, 25 hat keinen Modus.

Berechnen

Übung MMO-06

Modus

Bestimmen Sie das arithmetische Mittel $\overline{x}$, den Median $\tilde{x}_{0,5}$ und den Modus D der Zahlen der folgenden Reihe: 7, 6, 1, 3, 9, 9, 7, 4, 2, 7.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-07

Modus und Median

Bestimmen Sie das arithmetische Mittel $\overline{x}$, den Median $\tilde{x}_{0,5}$ und den Modus D der Zahlen der folgenden Reihe: 21, 6, 33, 3, 10, 8, 15, 4, 2, 14, 7, 20, 30, 1, 27.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

5 Unimodale bzw. multimodale Verteilungen und Schiefe (Exkurs)

Die Verteilung von Variablen ist in der Statistik von grundlegender Bedeutung. Das können Mietpreise, Körpergrößen, Einkommen und andere statistische Variablen sein. Neben anderen Charakteristika sind die Anzahl der Gipfel und die Symmetrie bzw. die Schiefe der Verteilung wichtig.

Wir verwenden hier ohne weitere Erklärung die **Dichtefunktion**. Sie zeigt ähnlich wie das Histogramm die Verteilung der Daten, allerdings für **stetige Variablen** deren Werte nicht klassiert wurden. Genauer werden Sie die Dichtefunktion in der Lerneinheit *„STV - Stetige Verteilungen“* kennenlernen.

Unimodale und multimodale Verteilung

Hat eine Dichtekurve nur ein lokales Maximum, dann spricht man von einer unimodalen Verteilung.

Hat eine Dichtekurve mehrere lokale Maxima, dann spricht man von einer bimodalen, trimodalen, etc. oder multimodalen Verteilung.

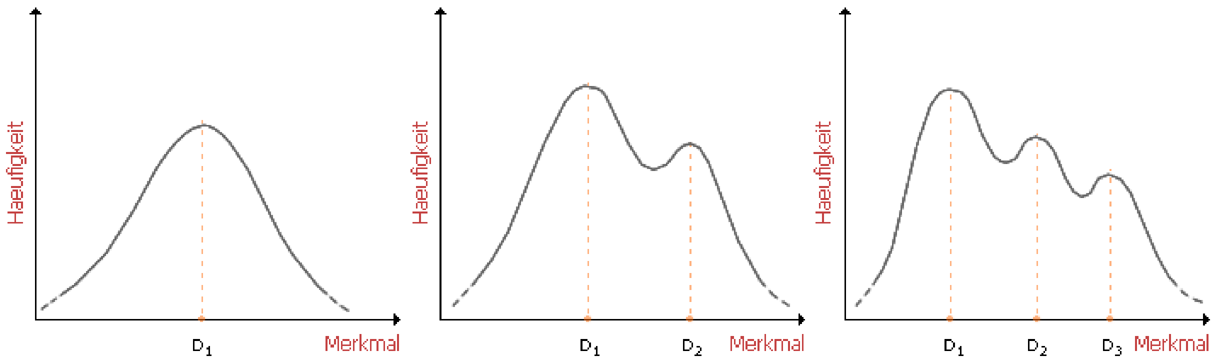


Abb.: Unimodale, bimodale, trimodale Verteilung mit dem eingezeichneten Modus D_i.

Symmetrie und Schiefe von Verteilungen

Ob eine Verteilung symmetrisch, links- oder rechtsschief ist, zeigt sich in der grafischen Darstellung der Dichtefunktion recht deutlich. Anhand von Modus, Median und arithmetischem Mittelwert lässt sich eine eingipflige Verteilung genauer charakterisieren.

$D < \tilde{x} < \overline{x}$ rechtsschief bzw. linkssteil
$D = \tilde{x} = \overline{x}$ symmetrisch
$\overline{x} < \tilde{x} < D$ linksschief bzw. rechtssteil.

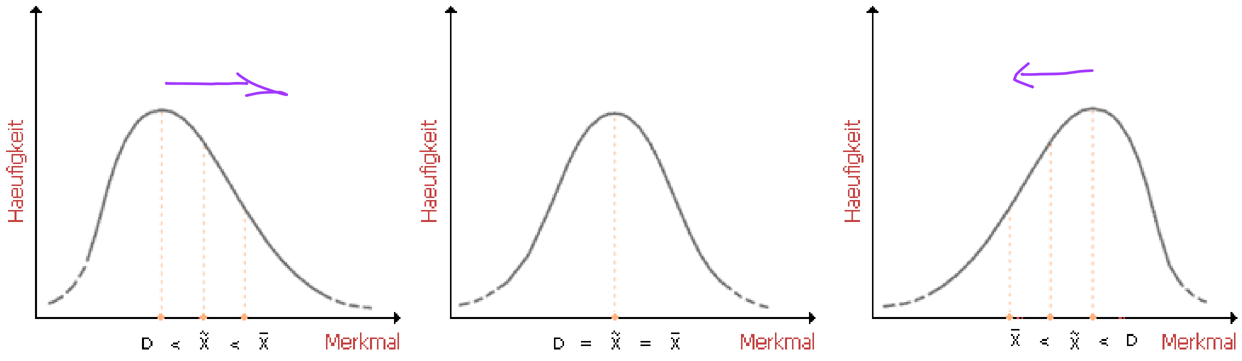


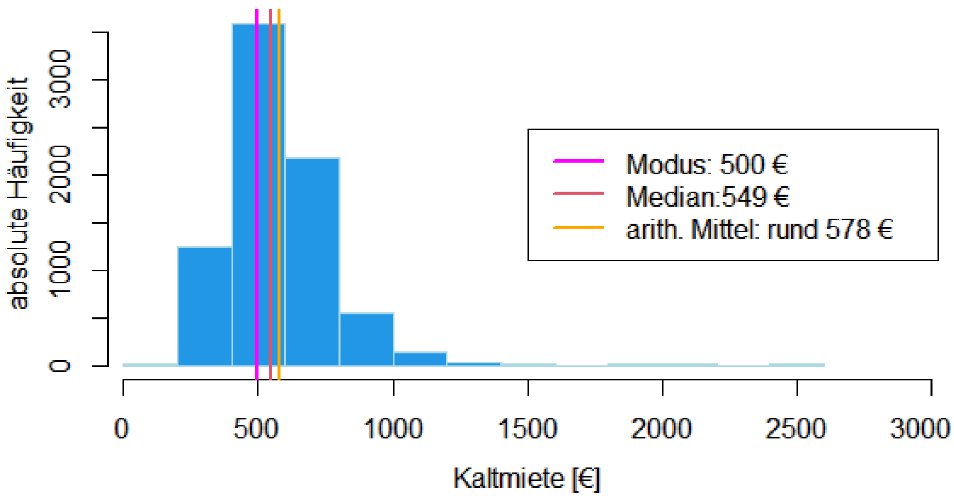
Abb.: Rechtsschiefe, symmetrische, linksschiefe Verteilungen mit eingezeichneten Lageparametern (Mittelwert $\overline{x}$, Median $\tilde{x}$, Modus D).

Die Schiefe ist zudem eine statistische Maßzahl deren Berechnung wir hier nicht ausführen. Sie finden sie aber in der gängigen Literatur.

Beispiel

Mieten in Potsdam für Zweizimmerwohnungen

Wir schauen uns die Mieten aus den Jahren 2014-2020 für Zweizimmerwohnungen in Potsdam an. Dafür nutzen wir den Ihnen schon bekannten Datensatz Potsdammieten.



Wie schön zu erkennen ist, handelt es sich um eine rechtsschiefe Verteilung mit Modus < Median < arithmetischem Mittel. Der Schwerpunkt bzw. die Mehrzahl der Mieten liegt klar im Intervall zwischen 400 und 600 € Kaltmiete. Zu den höheren Mieten hin nimmt die Häufigkeit nach und nach ab. Mieten jenseits von 1400 € sind nur noch vereinzelt vorhanden.

Wenn Sie das Beispiel einmal selber nachvollziehen wollen, finden Sie hier den R-Code und den Datensatz:

`Zweizimmerwohnung.R`
`Potsdammieten.csv` (2 KB)

6 Übungen zu Median und Modus

Berechnen

Übung MMO-08

Störfälle

In einem Weingeschäft in Venetto wurden in den Monaten Januar 2021 bis Dezember 2022 folgende Anzahlen von Störfällen registriert:

Monat	Jan	Feb	März	April	Mai	Juni	Juli	Aug	Sep	Okt	Nov	Dez
2021	0	4	1	4	2	8	2	4	3	2	1	6
2022	7	1	6	1	9	7	8	8	4	3	7	0

Bestimmen Sie den Median.

Stellen Sie den Erhebungsbefund in einem Säulendiagramm dar und geben Sie den Modus der Verteilung an.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-09

Häufigkeitsverteilung

Berechnen Sie den Median und den Modus für die folgende Häufigkeitsverteilung

Index i	x_i	Absolute Häufigkeit n_i
1	5	3
2	10	2
3	15	1
4	20	2

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-10

Wohnungsanzahl

Bezüglich der Anzahl der Wohnräume für ein bestimmtes Jahr in der Stadt Weibull ergab die Statistik des Wohnungsbestands folgendes Bild:

Raumanzahl	Wohnungsanzahl
1	130
2	615
3	1855
4	2720
5	1147
6	383
7 oder mehr	120

Bestimmen Sie Median und Modus.

Lösung (Siehe Anhang)

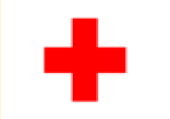
Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-11

Arbeitsunfälle

Im April des Jahres 2022 wurde die Anzahl der Arbeitsunfälle in der Firma Asli & Co durch die jeweils eine beschäftigte Person verletzt wurde, statistisch erfasst und aufbereitet. Es ergab sich, dass 90 % der Beschäftigten keinen Arbeitsunfall hatten, 7 % hatten einen Arbeitsunfall und 3 % hatten genau zwei Arbeitsunfälle.



Bestimmen und interpretieren Sie für diese Daten den Median und den Modus.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-12

Grundstücke

Selina, die Inhaberin des Weinfachgeschäftes Maestro, hat 10 neue Angebote von Grundstücken im Veneto bekommen. Die Grundstücke haben folgende Preise (in 10.000 €)

Angebot	1	2	3	4	5	6	7	8	9	10
Preis (in 10.000 €)	10,5	5	4,3	12	14	10,5	6,2	5	7,3	10,5

Bestimmen und interpretieren Sie den Median und den Modus der Preise von Angeboten.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-13

Verkaufte Flaschen

Die Inhaberin des Weinfachgeschäftes Maestro hat eine neue Sorte Prosecco in ihr Sortiment aufgenommen und interessiert sich für die Anzahl der pro Tag verkauften Flaschen dieser Sorte. Vier Wochen lang hat sie täglich die Anzahl der verkauften Flaschen notiert und in der folgenden Tabelle zusammengefasst:

Flaschenanzahl	Tage
0	3
2	5
3	7
5	4
7	3
8	2

Bestimmen und interpretieren Sie den Median und den Modus.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Zusammenfassung

- Der Median teilt die Daten in zwei Teile gleichen Umfangs, also in zwei Hälften.
- Der Median ist unempfindlich gegenüber Ausreißern in den Daten.
- Bei der Berechnung des Medians wird zwischen gerader und ungerader Anzahl von Datenpunkten unterschieden.
- Bei klassierten Daten wird der Median annähernd durch Interpolation bestimmt.
- Der Median minimiert die Summe der absoluten Abweichungen.
- Der Median ist mit monotonen Transformationen verträglich.
- Der Modus gibt den häufigsten Wert der Daten an.
- Der Modus kann bereits bei nominal skalierten Daten verwendet werden. (Siehe auch „*GST Kap. 5.5*“)

Sie sind am Ende dieser Lerneinheit angelangt. Auf den folgenden Seiten finden Sie noch die Übungen zur Wissensüberprüfung und die Übungen für die Statistiksoftware R.

Wissensüberprüfung

Mit den folgenden Übungen können Sie ihr Wissen überprüfen. Versuchen Sie zuerst, die Aufgaben selbst zu lösen, bevor Sie sich die Antworten anzeigen lassen.

Multiple Choice

Übung MMO-14

Welche Aussagen sind falsch und welche richtig?

	Richtig	Falsch	Auswertung
Der Median ist robuster gegenüber Ausreißern.			
Der Modus ist der Wert, der in einer nach der Größe geordneten Reihe $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ in der Mitte steht.			
Der Modus ist um so aussagefähiger, je stärker die betreffende Ausprägung dominiert.			
Der häufigste Wert ändert sich, wenn man die Merkmalsausprägungen umordnet.			

Multiple Choice

Übung MMO-15

Welche Aussagen sind falsch und welche richtig?

	Richtig	Falsch	Auswertung
Der Modus kann sich ändern, wenn man die Merkmalsausprägungen zusammenfasst.			
Der Median muss nicht unbedingt ein tatsächlich existierender Merkmalswert sein.			
Der Median ist weitgehend unabhängig von irgendwelchen extremen Werten.			
Je ausgefallener ein Wert ist, desto stärker wirkt er auf den Median ein.			

Lückentext

Übung MMO-16

Lückentext zu Median und Modus

Der Median ist bereits für _____-skalierte Daten geeignet während die Daten für das arithmetische Mittel mindestens _____-skaliert sein müssen. In die Berechnung des Medians gehen im Grunde nur ein oder zwei Werte direkt ein während beim arithmetischen Mittel _____ Datenwerte mit in die Berechnung einbezogen werden.

Der entscheidendste Unterschied liegt wohl in der _____. Während das arithmetische Mittel sehr _____ auf extreme Werte, Messfehler und andere Ausreißer reagiert, bleibt der Median davon _____. Neben der Berechnung des Median gibt es noch zwei _____ Möglichkeiten den Median zu bestimmen - die _____ und das _____.

Bei der Berechnung des Medians für _____ Daten ist es nicht von Interesse ob der Datenumfang gerade oder ungerade ist.

Der Modus ist der in den Daten am _____ vorkommende Wert. Gibt es keine gehäuft auftretenden Werte, wird der Modus als nicht _____ bezeichnet. Synonyme für Modus sind: _____ , _____ Wert, _____ Wert.

- alle
- dichtester
- empfindlich
- existent
- grafisch
- Histogramm
- häufigsten
- häufigster
- intervall
- klassiert
- Modalwert
- ordinal
- Robustheit
- unbeeinflusst
- Verteilungsfunktion

Übungen mit der Statistiksoftware R

Die in der Lerneinheit behandelten Themen können Sie anhand der folgenden Übungsaufgaben mit der Statistiksoftware **R** bearbeiten.

Installationshinweise [Manuals | **R** Installation and Administration]

Für einige Übungen stehen auch Musterlösungen für das Programm Excel bereit.

Berechnen

Übung MMO-17a

Häufigkeitsverteilung

Berechnen Sie den Median und den Modus aus der folgenden Häufigkeitsverteilung:

x_i	absolute Häufigkeiten
5	3
10	2
15	1
20	2

Lösung mit R und Excel (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Berechnen

Übung MMO-17b

Wohnungsbestand

Bezüglich der Anzahl der Wohnräume für ein bestimmtes Jahr in der Stadt N ergab die Statistik des Wohnungsbestands folgendes Bild:

Raumanzahl	Wohnungsanzahl
1	130
2	615
3	1855
4	2720
5	1147
6	383
7	120

Aufgabe

1. Bestimmen und interpretieren Sie den Median und den Modus.

Lösung mit R und Excel (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-17c

Störfälle

In einem Zulieferbetrieb wurden in den Monaten Januar 2021 bis Dezember 2022 folgende Anzahl von Störfällen registriert.

Monat	Jahr 2021	Jahr 2022
Jan	0	7
Feb	4	1
Mae	1	6
Apr	4	1
Mai	2	9
Jun	8	7
Jul	2	8
Aug	4	8
Sep	3	4
Okt	2	3
Nov	1	7
Dez	6	0

Aufgaben

1. Bestimmen Sie den Median
2. Stellen Sie den Erhebungsbefund in einem Säulendiagramm dar und geben Sie den Modus der Verteilung an.

Lösung mit R (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Berechnen

Übung MMO-17d

Grundstücke

Die Inhaberin des Weinfachgeschäftes Maestro hat 10 neue Angebote von Grundstücken in Veneto bekommen. Die Grundstücke haben folgende Preise (in 10.000 €):

Angebot	1	2	3	4	5	6	7	8	9	10
Preis (in 10.000 €)	10,5	5	4,3	12	14	10,5	6,2	5	7,3	10,5

Aufgabe

1. Bestimmen und interpretieren Sie den Median und den Modus der Preise von Angeboten.

Lösung mit R (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Berechnen

Übung MMO-17e

Prosecco

Die Inhaberin des Weinfachgeschäftes Maestro hat eine neue Sorte Prosecco in ihr Sortiment aufgenommen und interessiert sich für die Anzahl der pro Tag verkauften Flaschen dieser Sorte. Vier Wochen lang hat sie täglich die Anzahl der verkauften Flaschen notiert und in der folgenden Tabelle zusammengefasst.

Anzahl Flaschen	0	2	3	5	7	8
Anzahl Tage	3	5	7	4	3	2

Aufgabe

1. Bestimmen und interpretieren Sie den Median und den Modus.

Lösung mit R (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Berechnen

Übung MMO-17f

Wiederholungsprüfung

Die folgende Liste beinhaltet die Anzahl der Wiederholungsprüfungen im Fach Statistik von 117 Studierenden einer Berliner Hochschule, die im Verlauf des WS 2022 ihre Prufung absolvierten.

0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 2 0 0 0 0 0 0 0 1
0 0 0 1 0 0 0 0 0 1 0 0 1 0 0 3 0 0 1 0 0 0 0 2 0 1 0 1 0 1 0 0 0 1 0 0 1 1
0 1 0 0 1 0 0 0 0 0 1 0 0 0 1 0 2 1 1 0 0 0 0 0 1 0 1 1 1 1 0 2 0 1 1 2 1 1
0 1 0

Aufgabe

1. Charakterisieren Sie die Verteilung des Erhebungsmerkmals mit Hilfe von Median und Modus. Interpretieren Sie diese Verteilungsmaßzahl.

Lösung mit R (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Zusätzliche Übungsaufgaben

Berechnen

Übung MMO-18

Tagestemperaturen

Folgende Tagestemperaturen wurden in den Monaten Juni und Juli gemessen:

Juni	17	14,5	14	18	19,5	23	25					
Juli	23	18	19,5	25	23	26,5	25	27	28,5	26	25	28

(Alle Angaben in Grad Celsius)

Bestimmen Sie jeweils den Median für die Monate Juni und Juli.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-19

Körpergröße

Berechnen Sie den Median der folgenden Körpergrößen:

168, 168, 169, 170, 170, 171, 172, 172, 172, 172, 175, 175, 178, 179, 180, 181, 173, 174, 175

(Alle Angaben in Zentimeter)

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-20

Alter von Studierenden

die folgenden Liste zeigt das Alter von 25 Studierenden:

22, 20, 21, 23, 23, 24, 25, 28, 25, 23, 22, 21, 23, 24, 25, 21, 24, 23, 26, 24, 23, 25, 23, 21, 25

(Alle Angaben in Jahren)

Bestimmen Sie den Median.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-21

Versicherungsverträge

Ein Versicherungsmakler hat in 20 Tagen folgende Anzahl von Verträgen pro Tag abgeschlossen:

12, 15, 20, 23, 30, 13, 18, 17, 19, 25, 26, 14, 26, 19, 30, 32, 26, 28, 25, 20

Bestimmen Sie den Median.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-22

Tore der Fußball-Europameisterschaft

Während einer Fußball-Europameisterschaft fanden 31 Spiele statt. Im Folgenden finden Sie eine Liste mit der Anzahl der Tore je Spiel:

2, 5, 3, 2, 1, 1, 1, 1, 5, 3, 3, 3, 2, 4, 2, 4, 2, 1, 2, 3, 2, 5, 1, 2, 1, 6, 2, 6, 6, 3, 4

(Quelle: www.sportschau.de)

Bestimmen Sie den Median und Modus.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-23

Datenreihe 2

Bestimmen Sie das arithmetische Mittel, den Median und den Modus der Zahlen der folgenden Reihe:

23, 12, 24, 21, 26, 17, 29, 23, 19, 18, 21, 22, 23

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-24

Dummy-Werte

Bestimmen Sie für die folgenden Zahlen den Modus, die Quartile und den Median und erstellen Sie damit einen Boxplot:

21, 6, 33, 3, 10, 8, 15, 4, 2, 14, 7, 20, 30, 1, 27, 7, 2, 3, 4, 5,
10, 12, 17, 25, 6, 1, 3, 9, 9, 7, 4, 2, 7

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung MMO-25

Alter von Müttern bei Geburt ihres Kindes

Betrachten wir nochmal das Beispiel aus der Lerneinheit QXB.
Es liegt eine folgende Tabelle mit den Werten vor:

Alter	Anzahl der Geborenen
[15; 20)	70
[20; 25)	289
[25; 30)	369
[30; 35)	178
[35; 40)	83
[40; 45)	11

Geben Sie den Median und das durchschnittliche Alter der Mütter an und interpretieren Sie beide Werte vergleichend.

Lösung (Siehe Anhang)

Bearbeitungszeit: 8 Minuten

Berechnen

Übung MMO-26

Datenreihe 1

Bestimmen Sie für die folgende Datenreihe den Modus und den Median:

2, 2, 3, 3, 3, 5, 6, 6, 6, 7, 7, 8, 8, 9

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Appendix

Lösung für Übung MMO-01

Median bestimmen (1)

Man ordnet diese elf Werte zunächst der Größe nach:

2,5	3,5	3,7	4,1	4,5	5,1	5,7	6	7,1	7,5	7,7

$$\tilde{x} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{11+1}{2}\right)} = x_{\left(\frac{12}{2}\right)} = x_{(6)} = 5{,}1$$

Lösung für Übung MMO-02

Median bestimmen (2)

Man ordnet diese acht Klausurergebnisse zunächst der Größe nach:
67, 74, 78, 79, 87, 91, 92, 97.

Die Anzahl der Klausurergebnisse ist gerade. Daraus ergibt sich folgende Berechnung des Medians

$$\tilde{x} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) = \frac{1}{2}(79 + 87) = 83$$

Der Median der Punktzahlen liegt bei 83

Lösung für Übung MMO-03

Median bestimmen (3)

Die Gehälter [€] sind in einer Reihe geordnet:

1500, 1800, 1900, 2000, 2100, 2200, 2400, 5500.

Da die Anzahl der Gehälter gerade ist, gibt es zwei Werte in der Mitte, 2000 und 2100, deren Mittelwert ist gleich 2050.

$$\frac{1}{2}(2000 + 2100) = 2050$$

Somit ist der Median der Gehälter 2050 €.

Das arithmetische Mittel, das wir in Übung ARM-03 (Lerneinheit ARM, Kap. 2.2) berechnet haben lag bei 2425 €. Der Median wird durch den extremen Wert 5500 € nicht beeinflusst, während das beim arithmetischen Mittel durchaus der Fall ist. Dadurch charakterisiert der Median bei diesen Daten die durchschnittlichen Gehälter besser als das arithmetische Mittel.

Lösung für Übung MMO-04

Median bestimmen (4)

Man ordnet die Gehälter [€] in Abteilungen L, M, N und für alle Angestellten zusammen nach ihrer Größe:

- Abteilung L, n = 5: 1550, 1600, 1850, 2100, 2200.
- Abteilung M, n = 7: 1500, 1700, 1950, 2100, 2250, 2300, 2500.
- Abteilung N, n = 3: 1875, 2350, 2400.
- Abteilungen L, M und N zusammen, n = 15: 1500, 1550, 1600, 1700, 1850, 1875, 1950, 2100, 2100, 2200, 2250, 2300, 2350, 2400, 2500.

$$\tilde{x}_{(L)} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{5+1}{2}\right)} = x_{\left(\frac{6}{2}\right)} = x_{(3)} = 1850$$

$$\tilde{x}_{(M)} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{7+1}{2}\right)} = x_{\left(\frac{8}{2}\right)} = x_{(4)} = 2100$$

$$\tilde{x}_{(N)} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{3+1}{2}\right)} = x_{\left(\frac{4}{2}\right)} = x_{(2)} = 2350$$

$$\tilde{x}_{(L+M+N)} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{15+1}{2}\right)} = x_{\left(\frac{16}{2}\right)} = x_{(8)} = 2100$$

Der Median für die Abteilung L beträgt 1.850 €, für Abteilung M 2.100 € und für Abteilung N 2.350 €. ist einzeln und zusammen: Der Median für alle Angestellten der drei Abteilungen L, M und N zusammen beträgt 2.100 €.

Hinweis

Vorsicht: Anders als beim arithmetischen Mittel ist es nicht möglich den Median von zusammengefassten Daten aus den einzelnen Medianen zu berechnen.

Lösung für Übung MMO-05

Median bestimmen (5)

Man ordnet die sechs Werte zunächst der Größe nach:

$$x_{(1)} < x_{(2)} < x_{(3)} < x_{(4)} < x_{(5)} < x_{(6)}$$

Bei einer geraden Anzahl von Werten n = 6 erhalten wir den Median wie folgt:

$$\tilde{x} = \frac{1}{2}\left(x_{(\frac{n}{2})} + x_{((\frac{n}{2})+1)}\right) = \frac{1}{2}\left(x_{(\frac{6}{2})} + x_{((\frac{6}{2})+1)}\right) = \frac{1}{2}\left(x_{(3)} + x_{(3+1)}\right) = \frac{1}{2}\left(x_{(3)} + x_{(4)}\right)$$

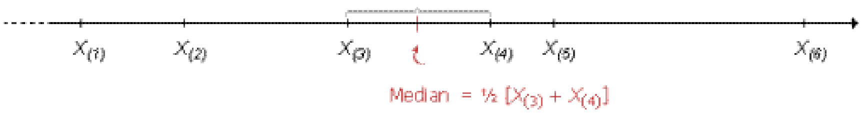


Abb.: Grafische Darstellung des Medians bei einer geraden Anzahl von Werten.

Lösung für Übung MMO-06

Das arithmetische Mittel

$$\overline{x} = \frac{7+6+1+3+9+9+7+4+2+7}{10} = 5{,}5$$

Der Median:
Die Zahlen geordnet: 1, 2, 3, 4, 6, 7, 7, 7, 9, 9

n ist gerade, nach der Formel

$$Z = \tilde{x}_{0{,}5} = \begin{cases} x_{\left(\frac{(n+1)}{2}\right)} & \text{falls } n \text{ ungerade} \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\left(\frac{n}{2}\right)+1\right)}\right) & \text{falls } n \text{ gerade} \end{cases}$$

gilt dann

$$\tilde{x}_{0{,}5} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\left(\frac{n}{2}\right)+1\right)}\right) = \frac{1}{2}\left(x_{\left(\frac{10}{2}\right)} + x_{\left(\left(\frac{10}{2}\right)+1\right)}\right) = \frac{1}{2}\left(x_{5)} + x_{(6)}\right)$$

$$= \frac{1}{2}(6+7) = \frac{1}{2} \cdot 13 = 6{,}5$$

Der Modus:
Die am häufigsten vorkommende Zahl ist: D = 7

Lösung für Übung MMO-07

Das arithmetische Mittel

$$\overline{x} = \frac{21+6+33+3+10+8+15+4+2+14+7+20+30+1+27}{15} = 13{,}4$$

Der Median:
Die Zahlen geordnet: 1, 2, 3, 4, 6, 7, 8, 10, 14, 15, 20, 21, 27, 30, 33

n ist ungerade, nach der Formel

$$Z = \tilde{x}_{0{,}5} = \begin{cases} x_{\left(\frac{(n+1)}{2}\right)} & \text{falls } n \text{ ungerade} \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\left(\frac{n}{2}\right)+1\right)}\right) & \text{falls } n \text{ gerade} \end{cases}$$

gilt dann

(oder die Zahl in der Mitte):

$$\tilde{x}_{0{,}5} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{15+1}{2}\right)} = x_{\left(\frac{16}{2}\right)} = x_{(8)} = 10$$

Der Modus:
Alle Werte kommen genau einmal vor. Es existiert kein Modus

Lösung für Übung MMO-08

Störfälle

Man ordnet die 24 Werte zunächst der Größe nach an:
0 0 1 1 1 1 2 2 2 3 3 4 4 4 4 6 6 7 7 7 8 8 8 9

Mit n = 24 ist der Median in:

$$\tilde{x} = \frac{1}{2}\left(x_{(\frac{n}{2})} + x_{((\frac{n}{2})+1)}\right) = \frac{1}{2}\left(x_{(\frac{24}{2})} + x_{((\frac{24}{2})+1)}\right) = \frac{1}{2}\left(x_{12)} + x_{(12+1)}\right)$$
$$= \frac{1}{2}\left(x_{(12)} + x_{(13)}\right) = \frac{1}{2}(4+4) = 4$$

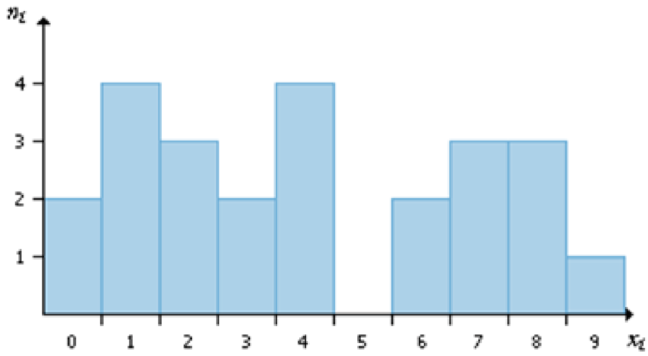


Abbildung: Anzahl der Störfälle im Monat

Der Modus ist nicht eindeutig.

Am häufigsten werden X = 1 und X = 4 Störfälle (jeweils viermal) beobachtet.

Lösung für Übung MMO-09

Häufigkeitsverteilung

Median:

Man ordnet diese acht Werte zunächst der Größe nach: 5, 5, 5, 10, 10, 15, 20, 20.

Dann ist der Median $\tilde{x}$, hier also:

$$\tilde{x} = \frac{1}{2}\left(x_{(\frac{n}{2})} + x_{((\frac{n}{2})+1)}\right) = \frac{1}{2}\left(x_{(\frac{8}{2})} + x_{((\frac{8}{2})+1)}\right) = \frac{1}{2}\left(x_{4)} + x_{(4+1)}\right)$$
$$= \frac{1}{2}\left(x_{(4)} + x_{(5)}\right) = \frac{1}{2}(10+10) = 10$$

Mindestens die Hälfte der Beobachtungen sind 10 oder mehr.

Modus:

D = 5

Die Zahl 5 kommt bei allen Beobachtungen am häufigsten vor.

Lösung für Übung MMO-10

Wohnungsanzahl

Es gibt insgesamt n=6970 Wohnungen (Addieren der Häufigkeiten)

Median:

$$\tilde{x} = x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} = x_{(\frac{6970}{2})} = x_{(\frac{6971}{2})} = 4$$

Die Hälfte aller Wohnungen hat 4 oder mehr Zimmer als auch 4 oder weniger Zimmer, deshalb muss der Median bei 4 liegen.

Modus:

Der Modus ist offensichtlich auch D=4, da am häufigsten (2720 mal) Vierzimmerwohnungen angeboten werden.

Lösung für Übung MMO-11

Arbeitsunfälle

Median:

$$\tilde{x} = 0$$

0 Arbeitsunfälle je Beschäftigte: d. h. mindestens 50 % der Beschäftigten hatten keinen Arbeitsunfall.

Modus:

D = 0
Arbeitsunfälle je Beschäftigte: d. h. 0 Arbeitsunfälle je Beschäftigte sind am häufigsten.

Lösung für Übung MMO-12

Grundstücke

Median:

Man ordnet diese zehn Werte zunächst der Größe nach an:

4,3	5	5	6,2	7,3	10,5	10,5	10,5	12	14

$$\begin{aligned}\tilde{x} &= \frac{1}{2}\left(x_{(\frac{n}{2})} + x_{((\frac{n}{2})+1)}\right) = \frac{1}{2}\left(x_{(\frac{10}{2})} + x_{((\frac{10}{2})+1)}\right) = \frac{1}{2}\left(x_{5)} + x_{(5+1)}\right)\\ &= \frac{1}{2}\left(x_{(5)} + x_{(6)}\right) = \frac{1}{2}(7{,}3 + 10{,}5) = 8{,}9\end{aligned}$$

D. h. die eine Hälfte der Grundstücke kostet weniger als 89.000 und die andere Hälfte mehr.

Modus:

D = 10,5: d. h. Angebote in Höhe von 105.000 Euro sind am häufigsten.

Lösung für Übung MMO-13

Verkaufte Flaschen

Median:

Die geordnete Anzahl verkaufter Flaschen sieht wie folgt aus:

0 0 0 2 2 2 2 2 3 3 3 3 3 3 3 5 5 5 5 7 7 7 8 8.

Somit gilt n=24 und der Median liegt zwischen dem 12. und 13. größten Wert. Ein scharfer Blick, und wir sehen $\tilde{x} = 3$.

Modus:

D = 3 Flaschen, d. h. am häufigsten wurden 3 Flaschen pro Tag verkauft. Der Modalwert ist hier aussagekräftiger.

Lösung für Übung MMO-17a

Häufigkeitsverteilung

1. Der Median ist 10
2. Der Modus ist direkt aus der Häufigkeitsverteilung ablesbar: Modus = 5

xi	absolute Häufigkeiten
5	3
10	2
15	1
20	2

Lösung mit R

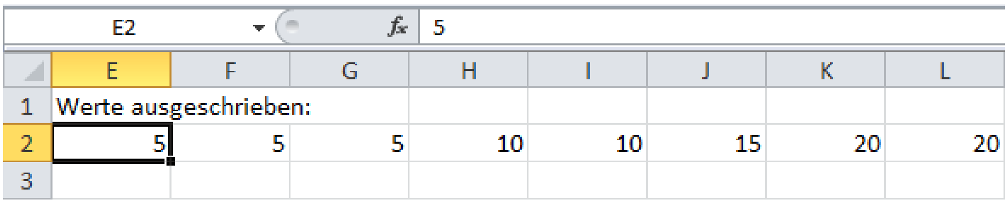
hverteilung_loesung.R

```
xi = c(5, 10, 15, 20) # Vektor der Werte
N = c(3, 2, 1, 2) # Vektor der absoluten Häufigkeiten
data=rep(xi,N) # rep Wiederholt die Werte entsprechend ihrer Häufigkeit
data # Ausgabe der geordneten Werte
median(data) # Median der Werte
```

Lösung mit Excel

WMS_MMO_17a_Haeufigkeitsverteilung.xlsx (10 KB)

Um Median und Modus zu berechnen, schreiben wir die Daten zunächst in folgender Form auf:



Dann berechnen wir mit den Funktionen `MEDIAN` und `MODUS.EINF` Median und Modus. Wie das genau geht, ist in den folgenden Abbildungen zu sehen.

B11 f_x =MEDIAN(E2:L2)

	A	B	C	D	E
8	Aufgabe				
9					
10	Median:				
11		10			

B13 f_x =MODUS.EINF(E2:L2)

	A	B	C	D	E
12	Modus:				
13		5			
14					

Alternativ können wir Modus und Median durch Abzählen berechnen. Das geht so:

	A	B	C
15			
16	alternative Lösung:		
17	n=8		
18	n/2=4		
19	Der vierte Wert ist eine 10.		
20	Der fünfte Wert ist auch eine 10.		

So erhalten wir die folgende Lösung:

Median:	10		
Modus:	5	(direkt ablesbar)	

Lösung für Übung MMO-17b

Wohnungsbestand

1. Der Median ist 4
2. Der Modus ist direkt aus der Häufigkeitsverteilung ablesbar: Modus = 4

Interpretation: Da es sich bei dem Wohnungsbestand um eine relativ symmetrische Verteilung handelt, ist der Median gleich dem Modus. Die meisten Wohnungen haben 4 Zimmer und da die Verteilung des Wohnungsbestandes links und rechts vom Modus gleichermaßen abflacht, ist auch der Median gleich 4.

Lösung mit R

wbestand_loesung.R

```
Raeume=c(1:7)  # Ein Vektor von 1,..,7 (Zimmeranzahl)
N = c(130, 615, 1855, 2720, 1147, 383, 120)  # Vektor der absoluten Häufigkeiten
data=rep(Raeume,N)  # rep Wiederholt Raumanzahl entsprechend ihrer absoluten Häufigkeit
data # Ausgabe der geordneten Werte
median(data) # Median der Werte
```

Lösung mit Excel

WMS_MMO_17b_Wohnungsbestand.xlsx (9 KB)

***Aufgabe 1**: Bestimmen und interpretieren Sie den Median und den Modus.*

In dieser Aufgabe ist es in Excel zu mühsam, die Daten in anderer Form aufzuschreiben. Deshalb berechnen wir den Median hier nur mit der aus der Aufgabe Häufigkeitsverteilung bekannten alternativen Methode. Das geht so, wie es im folgenden Bild dargestellt ist:

	A	B	C
10			
11	Aufgabe		
12			
13	Median:		
14	n:	6970	
15	n/2:	3485	
16	Der 3485. Wert ist eine 4.		
17	Der 3486. Wert ist auch eine 4.		
18	Der Median ist 4.		
19			
20	Modus:		
21	Direkt abzulesen:		
22		4	

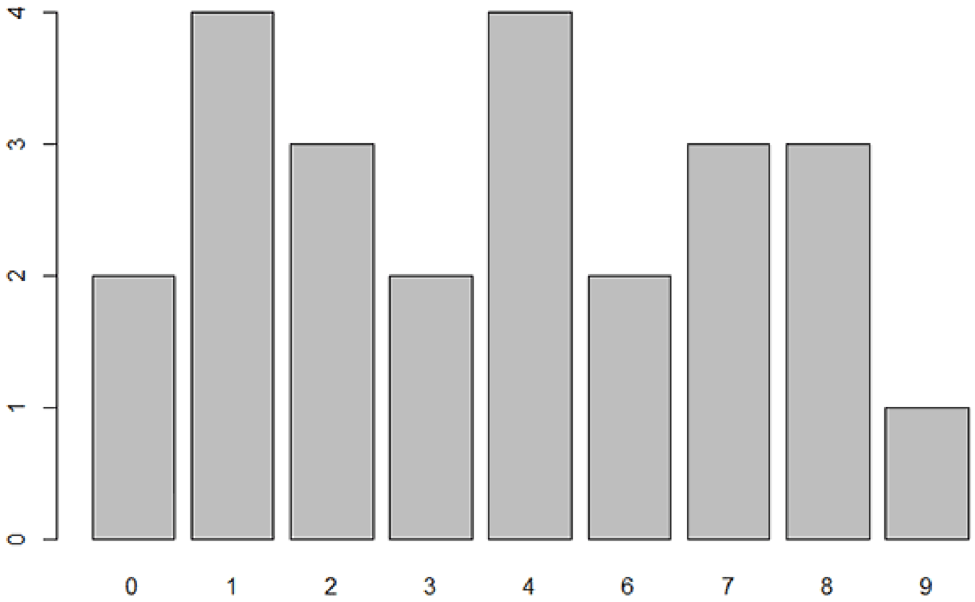
So kommen wir zu folgendem Ergebnis:

Median:	4
Modus:	4

Lösung für Übung MMO-17c

Störfälle

1. Der Median ist 4.
2. Das Säulendiagramm zeigt den Erhebungsbefund. Die Verteilung der Störfälle besitzt zwei Modi. Die häufigsten Störfälle pro Monat waren 1 und 4 und das sind somit auch die Modi der Verteilung.



Lösung mit R

stoerfaelle_loesung.R

```
stoerfaelle <-
  data.frame(
    Monat = c(
      "Jan",
      "Feb",
      "Mae",
      "Apr",
      "Mai",
      "Jun",
      "Jul",
      "Aug",
      "Sep",
      "Okt",
      "Nov",
      "Dez"
    ),
    Jahr_2021 = c(0, 4, 1, 4, 2, 8, 2, 4, 3, 2, 1, 6),
    Jahr_2022 = c(7, 1, 6, 1, 9, 7, 8, 8, 4, 3, 7, 0)
  )

median(c(stoerfaelle$Jahr_2021,stoerfaelle$Jahr_2022))

barplot(table(c(stoerfaelle$Jahr_2021,stoerfaelle$Jahr_2022)))
```

Lösung für Übung MMO-17d

Grundstücke

1. Der Median liegt bei 8.9. Der Modus ist an der Häufigkeitstabelle ablesbar table(angebote). Der häufigste Wert und somit der Modus ist 10,5.

4,3	5	6,2	7,3	10.5	12	14
1	2	1	1	3	1	1

Interpretation: Aufgrund der fünf kleineren Angebote als der Modus ist der Median auch kleiner als der häufigste Wert.

Lösung mit R

grundstuecke_loesung.R

```
angebote<-c(10.5,5,4.3,12,14,10.5,6.2,5,7.3,10.5)

median(angebote)

table(angebote)
```

Lösung für Übung MMO-17e

Prosecco

1. Der Median ist 3. Der Modus ist direkt aus der Häufigkeitsverteilung ablesbar: Modus = 3

Anzahl Flaschen	0	2	3	5	7	8
Anzahl Tage	3	5	7	4	3	2

Interpretation: Links und rechts vom Modus sind ungefähr gleich viele Häufigkeiten aufgetreten. Daraus resultiert, dass der Median gleich dem Modus ist.

Lösung mit R

prosecco_loesung.R

```
Flaschen= c(0, 2, 3, 5, 7, 8)    # Vektor der verkauften Flaschenanzahl pro Tag
Tage = c(3, 5, 7, 4, 3, 2)     # Vektor der Tage
data=rep(Flaschen, Tage)  # rep wiederholt die Flaschenanzahl entsprechend der Tage
data # Ausgabe der geordneten Werte
median(data) # Median
```

Lösung für Übung MMO-17f

Wiederholungsprüfung

1. Der Median ist 0. Aus der Häufigkeitstabelle lässt sich ablesen, dass der Modus 0 ist.

Wiederholungsversuche	0	1	2	3
Anzahl Studierende	81	30	5	1

Interpretation: In diesem Beispiel herrscht ein großes Übergewicht an Studierenden, welche die Prüfung im ersten Anlauf geschafft haben. Deshalb ist sowohl Median als auch Modus gleich 0

Lösung mit R

wpruefung_loesung.R

```
absolventenliste <-
  c(
  0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,1,
  0,0,0,0,0,0,0,1,0,0,2,0,0,0,0,0,0,0,1,
  0,0,0,1,0,0,0,0,0,1,0,0,1,0,0,3,0,0,1,
  0,0,0,0,2,0,1,0,1,0,1,0,0,0,1,0,0,1,1,
  0,1,0,0,1,0,0,0,0,0,1,0,0,0,1,0,2,1,1,
  0,0,0,0,0,1,0,1,1,1,1,0,2,0,1,1,2,1,1,
  0,1,0
  )

median(absolventenliste)

table(absolventenliste)
```

Lösung für Übung MMO-18

Tagestemperaturen

Der Median für den Monat Juni ist 18, der für Juli ist 25.

Lösung mit R

```
daten1<-c(17, 14.5, 14, 18, 19.5, 23, 25)
daten2<-c(23, 18, 19.5, 25, 23, 26.5, 25, 27, 28.5, 26, 25, 28)
Juni<-sort(daten1)
Juli<-sort(daten2)

Juni
[1] 14.0 14.5 17.0 18.0 19.5 23.0 25.0

Juli
[1] 18.0 19.5 23.0 23.0 25.0 25.0 25.0 26.0 26.5 27.0 28.0 28.5

m_Juni<-median(Juni)
m_Juli<-median(Juli)

m_Juni
[1] 18

m_Juli
[1] 25
```

Lösung für Übung MMO-19

Körpergröße

Der Median beträgt 172 cm.

Lösung mit R

```
kgr<-
  c(168, 168, 169, 170, 170, 171, 172, 172, 172, 172,
    173, 174, 175, 175, 175, 178, 179, 180, 181)
m_kgr<-median(kgr)

m_kgr
[1] 172
```

Lösung für Übung MMO-20

Alter von Studierenden

Der Median beträgt 23.

Lösung mit R

```
alter<-
  c(22, 20, 21 ,23 ,23, 24, 25, 28, 25 ,23, 22, 21, 23,
    24, 25, 21, 24, 23 ,26 ,24, 23, 25, 23, 21, 25)

sort_alter<-sort(alter)
sort_alter

[1] 20 21 21 21 21 22 22 23 23 23 23 23 23 23 24 24 24 24 25 25 25 25 25 26 28

m_alter<-median(sort_alter)

m_alter
[1] 23
```

Lösung für Übung MMO-21

Versicherungsverträge

Der Median beträgt 21,5

Lösung mit R:

```
vert<-
        c(12, 15, 20, 23, 30, 13, 18, 17, 19, 25,
          26, 14, 26, 19, 30, 32, 26, 28, 25, 20)
sort_vert<-sort(vert)

sort_vert
[1] 12 13 14 15 17 18 19 19 20 20 23 25 25 26 26 26 28 30 30 32

m_vert<-median(sort_vert)

m_vert
[1] 21.5
```

Lösung für Übung MMO-22

Tore der Fußball-Europameisterschaft

Der Median und auch der Modus ist 2

Lösung mit R

```
tore<-
        c(2, 5, 3, 2, 1, 1, 1, 1, 5, 3, 3, 3, 2, 4, 2,
          4, 2, 1, 2, 3, 2, 5, 1, 2, 1, 6, 2, 6, 6, 3, 4)
sort_tore<-sort(tore)

sort_tore
[1] 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 4 4 4 5 5 5 6 6 6

m_tore<-median(sort_tore)

m_tore
[1] 2

table(tore)
tore
1 2 3 4 5 6
7 9 6 3 3 3
```

Lösung für Übung MMO-23

Datenreihe 2

Das arithmetische Mittel:

$$\overline{x} = \frac{23+12+24+21+26+17+29+23+19+18+21+22+23}{13} = 21{,}38$$

Der Median ist 22:

Der Modus kann aus der Häufigkeitstabelle abgelesen werden. Die am häufigsten vorkommende Zahl ist 23.

12	17	18	19	21	22	23	24	26	29
1	1	1	1	2	1	3	1	1	1

Lösung für R

```
 datenreihe<-c(23, 12, 24, 21, 26, 17, 29, 23, 19, 18, 21, 22, 23)
sort_drh<-sort(datenreihe)
sort_drh
median(sort_drh)

table(datenreihe)
datenreihe
```

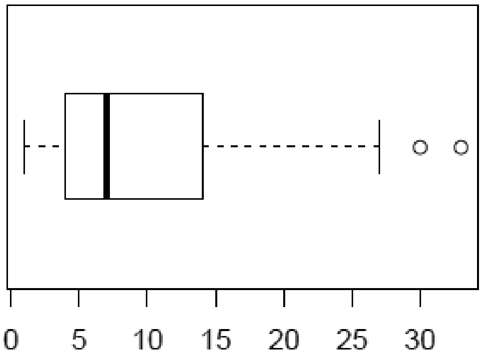
Lösung für Übung MMO-24

Dummy-Werte

Modus = 7
Median = 7
Quartile

0%	25%	50%	75%	100%
1	4	7	14	33

Als Minimum bekommen wir 1 und als Maximum bekommen wir 33. Die Quartile, die wir für den Boxplot brauchen, sind $Q_{0,25}$ = 4 und $Q_{0,75}$ = 14 und der Median ist gleich 7.



Lösung mit R

```
zahlen<-
        c(21, 6, 33, 3, 10, 8, 15, 4, 2, 14, 7, 20, 30, 1, 27, 7,
          2, 3, 4, 5, 10, 12, 17, 25, 6, 1, 3, 9, 9, 7, 4, 2, 7)
sort_zahlen<-sort(zahlen)

sort_zahlen
[1] 1 1 2 2 2 3 3 3 4 4 4 5 6 6 7 7 7 7 8 9 9 10 10 12 14
[26] 15 17 20 21 25 27 30 33

quantile(zahlen)
0% 25% 50% 75% 100%
1 4 7 14 33

table(zahlen)
zahlen
1 2 3 4 5 6 7 8 9 10 12 14 15 17 20 21 25 27 30 33
2 3 3 3 1 2 4 1 2 2 1 1 1 1 1 1 1 1 1 1

min(zahlen)
[1] 1

max(zahlen)
[1] 33

median(zahlen)
[1] 7

boxplot(sort_zahlen,horizontal=TRUE)
```

Lösung für Übung MMO-25

Alter von Müttern bei Geburt ihres Kindes

Bestimmung der relativen Häufigkeiten und relativen Summenäufigkeiten:

Alter	Anzahl der Geborenen	h_i	$F(x_i)$
[15; 20)	70	0,07	0,07
[20; 25)	289	0,289	0,359
[25; 30)	369	0,369	0,728
[30; 35)	178	0,178	0,906
[35; 40)	83	0,083	0,989
[40; 45)	11	0,011	1

Somit sind 50 % der Mütter nicht älter als 30.

Bestimmung des Median:

$$\text{Median } = 25 + \frac{0{,}5 - 0{,}359}{0{,}369}(30 - 25) = 26{,}91$$

Somit sind 50 % der Mütter höchstens 26,91 Jahre alt.

Bestimmung der Klassenmitten:

Alter	Anzahl der Geborenen	x_iM
[15; 20)	70	17,5
[20; 25)	289	22,5
[25; 30)	369	27,5
[30; 35)	178	32,5
[35; 40)	83	37,5
[40; 45)	11	42,5

Bestimmung des durchschnittlichen Alters der Mütter:

$$\overline{x} = \frac{1}{1000}(17{,}5 \cdot 70 + 22{,}5 \cdot 289 + 27{,}5 \cdot 369 + 32{,}5 \cdot 178 + 37{,}5 \cdot 83 + 42{,}5 \cdot 11) = 27{,}24$$

Somit sind die Mütter im Durchschnitt 27,24 Jahre alt.

Interpretation:50 % der Mütter sind höchstens 27 Jahre alt und im Durchschnitt haben sie ihre Kinder mit 27 Jahren zur Welt gebracht.

Lösung für Übung MMO-26

Datenreihe 1

Die Datenreihe ist bimodal und hat zwei Modi 3 und 6 und als Median der Daten bekommen wir 6.