Stochastik

Mit dem Begriff **Stochastik** werden in der Mathematik die Bereiche Wahrscheinlichkeitstheorie und Statistik zusammengefasst. Die „Kunst des Vermutens" – wie sich Stochastik übersetzen lässt – beschäftigt sich mit Ereignissen und Ergebnissen, die unterschiedlich häufig auftreten, wenn sich ein Vorgang wiederholt.

Man möchte:

- den Ausgang von Zufallsexperimenten abschätzen
- statistischeVerteilungen beschreiben und beurteilen

1 Unterschied Stochastik zu Statistik

Der Unterschied zwischen Stochastik und Statistik ist relativ simpel. Die Statistik ist neben der Wahrscheinlichkeitsrechnung ein Teilbereich der Stochastik. Die Statistik setzt sich mit Datenmengen auseinander und betrachtet und interpretiert die Verteilungen dieser Daten.

2 Wahrscheinlichkeitsrechnung

In der Wahrscheinlichkeitslehre verwendest du die ein oder andere Formel, um Wahrscheinlichkeiten zu berechnen.

Es ist wichtig, zu verstehen, dass das tatsächliche Eintreffen eines einzelnen Ereignisses oder weniger Ereignisse nicht durch Wahrscheinlichkeitsrechnung vorhergesagt werden kann: Wenn Sie etwa sechsmal hintereinander würfeln, wird es sehr selten vorkommen, dass jede der Zahlen 1 bis 6 dabei herauskommt. Würfeln Sie dagegen sechstausendmal, ist es wahrscheinlich, dass die Häufigkeit jeder der sechs Zahlen nicht allzu weit von 1000 entfernt liegt (falls doch, stimmt etwas mit Ihrem Würfel nicht).

Ein Ereignis, das nie eintritt, hat die Wahrscheinlichkeit 0, eines, das bei jedem Versuch eintritt, hat die Wahrscheinlichkeit 1, und alle anderen Fälle werden als Brüche zwischen 0 und 1 ausgedrückt. Genauer gesagt wird die Anzahl der erwünschten Ereignisse durch die Gesamtzahl der Ereignisse geteilt, sodass es sich bei Wahrscheinlichkeiten stets um rationale Zahlen handelt. Alternativ werden Wahrscheinlichkeiten in Prozent angegeben; eine Wahrscheinlichkeit von ist beispielsweise dasselbe wie 50 %.

2.1 Definition:

In der Wahrscheinlichkeitsrechnung ermittelst du Wahrscheinlichkeiten für bestimmte Ereignisse in einem Zufallsexperiment, zum Beispiel die Wahrscheinlichkeit für Kopf beim Werfen einer Münze.

Die Wahrscheinlichkeitsrechnung beschäftigt sich mit:

- Zufallsexperimenten
- Ergebnisse
- Ereignisse
- La Place Experimente

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

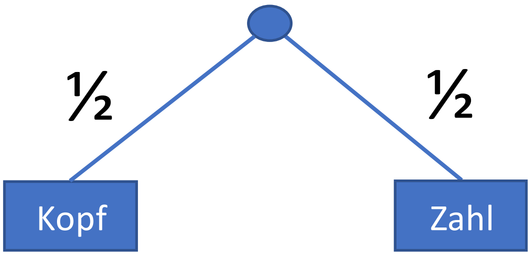
2.2 Beispiel:

Wenn du zum Beispiel eine Münze wirfst, erhältst du zufällig Kopf oder Zahl. Du weißt vorher nicht, was herauskommt. Aber du kannst berechnen, wie groß die Wahrscheinlichkeit für ein Ergebnis ist (z.B. für Kopf). Damit beschäftigt sich die **Wahrscheinlichkeitsrechnung**.

In unserem Beispiel sind Kopf und Zahl beide gleich wahrscheinlich. Du hast also eine Fifty-Fifty-Chance: Die Wahrscheinlichkeit für Kopf ist ½ (oder 50 %) und die für Zahl auch.

Baumdiagramm der Wahrscheinlichkeitsrechnung

Das Zufallsexperiment mit der Münze kannst du an einem Baumdiagramm darstellen. Dabei zeichnest du ganz oben einen Knoten und dann zwei Linien nach unten, weil es zwei mögliche Ergebnisse gibt (Kopf und Zahl). Du nennst diese Linien Pfade. Neben sie schreibst du die Wahrscheinlichkeiten für das jeweilige Ergebnis, hier also ½.



An einem Baumdiagramm, auch Ereignisbaum genannt, kannst du ganz leicht verstehen, wie dein Zufallsexperiment aufgebaut ist. Das hilft dir bei der Wahrscheinlichkeitsrechnung enorm!

2.3 Laplace-Experiment

In der Wahrscheinlichkeitsrechnung bedeutet Fairness, dass die Wahrscheinlichkeit für jedes gleichartige Ereignis identisch ist. Seit Jahrtausenden versuchen Menschen, die Glücksspiele anbieten oder spielen, die Gewinnwahrscheinlichkeit jeweils zu ihren Gunsten zu verschieben, indem sie beispielsweise Würfel mit ungleicher Massenverteilung bauen. Faire Experimente mit endlich vielen möglichen Ergebnissen werden als Laplace-Experimente bezeichnet.

Wenn du wie im vorherigen Beispiel eine Münze wirfst, dann sind Kopf und Zahl gleich wahrscheinlich. Beide Ergebnisse haben eine Chance von ½. Wenn alle möglichen Ergebnisse die gleiche Wahrscheinlichkeit haben, sprichst du von einem Laplace-Experiment .

Beispiel:

Du wirfst einen Würfel. Jede Zahl zwischen 1 und 6 ist dann gleich wahrscheinlich. Es handelt sich also um ein **Laplace-Experiment**. Die Chance für jede Zahl liegt bei 1/6.

Nicht alle Experimente sind Laplace-Experimente. Stell dir vor, in einer Lostrommel sind 2 Gewinne und 8 Nieten und du ziehst ein Los. Dann ist es wahrscheinlicher, dass du eine Niete ziehst. Die beiden Ergebnisse haben also nicht die gleiche Wahrscheinlichkeit.

Bei Laplace-Experimenten kannst du die Wahrscheinlichkeit mit einer Formel berechnen. Dabei teilst du die Anzahl der betrachteten Ergebnisse durch die Anzahl aller möglichen Ergebnisse (beim Würfel 6 mögliche Ergebnisse, nämlich 1 bis 6). Die Wahrscheinlichkeit, eine 3 oder eine 5 (also 2 betrachtete Ergebnisse) zu würfeln ist also:

$$\frac{\text{Betrachtete Ergebnisse}}{\text{Alle möglichen Ergebnisse}} = \frac{2}{6} = \frac{1}{3}$$

Oder:

$$P(3) + P(5) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

2.4 Zufallsexperimente

Die Münze, der Würfel und die Lose — all das sind **Zufallsexperimente**. Das bedeutet, dass es mindestens zwei mögliche Ausgänge gibt und du vorher nicht weißt, was herauskommen wird.

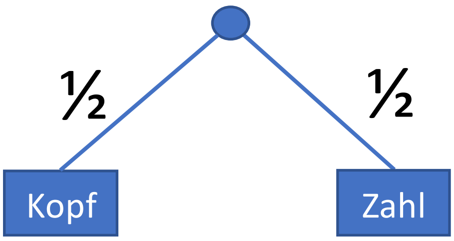
Du unterscheidest zwei verschiedene Arten von Zufallsexperimenten: einstufige und mehrstufige.

Bisher hast du nur einstufige Zufallsexperimente betrachtet. Das heißt, dass du deinen Versuch nur einmal durchgeführt hast. Zum Beispiel wurde die Münze einmal geworfen.

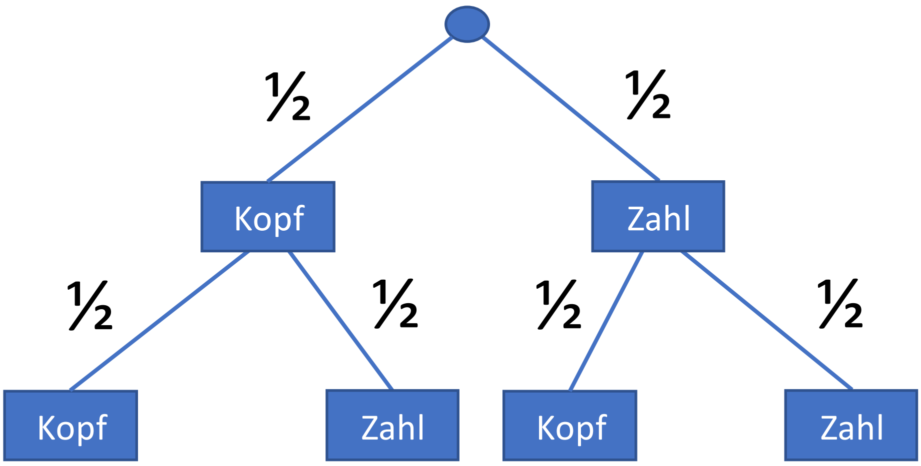
Führst du Experimente aber öfter durch. Dann sprichst du von einem mehrstufigen Zufallsexperiment.

Du kannst auch mehrstufige Zufallsexperimente in ein Baumdiagramm einzeichnen. Es hat dann so viele Ebenen, wie du Versuchsdurchgänge machst:

Einstufiges Zufallsexperiment



mehrstufiges Zufallsexperiment



2.4.1 Berechnung mehrstufiger Wahrscheinlichkeiten

Soll dagegen die kombinierte Wahrscheinlichkeit mehrerer einzelner Ereignisse berechnet werden, gelingt dies durch Multiplikation der Wahrscheinlichkeiten. Wenn etwa zwei Münzen (oder zwei aufeinanderfolgende Würfe derselben Münze) Kopf zeigen sollen, wird dies so berechnet:

$$P(\text{Kopf}) * P(\text{Kopf}) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

Manchmal ist es am praktischsten, die Wahrscheinlichkeit einer bestimmten Ereigniskombination zu berechnen, indem die Wahrscheinlichkeit des Gegenereignisses von 1 abgezogen wird. Das Gegenereignis ist das Gegenteil eines Ereignisses. Wie groß ist beispielsweise die Wahrscheinlichkeit, bei zwei Münzwürfen mindestens einmal Kopf zu werfen? Das Ereignis »mindestens einmal Kopf« ist das Gegenereignis von »zweimal Zahl«, also gilt

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

P(Kopf >= 1) = 1 – (P(Zahl) * P(Zahl)) = $1 - (\frac{1}{2} * \frac{1}{2}) = \frac{3}{4}$

2.5 Abhängige und unabhängige Ereignisse

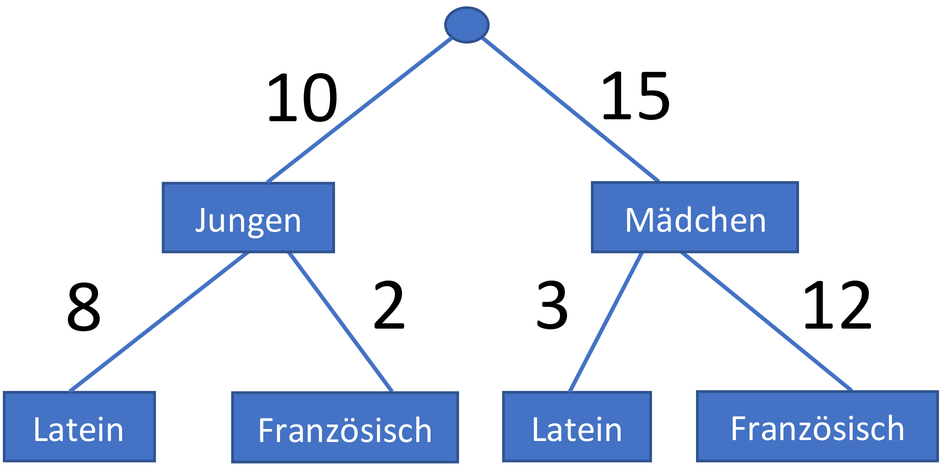
Du kennst jetzt schon mehrstufige Zufallsexperimente, bei denen alle Ebenen im Baumdiagramm genau gleich aussehen. Das ist aber nicht immer der Fall. Schau dir zum Beispiel folgendes Zufallsexperiment an:

Aus einer Klasse wird zufällig ein Schüler ausgewählt. In der Klasse sind 10 Jungen und 15 Mädchen.

Von den Jungen lernen 8 Latein und 2 Französisch.

Von den Mädchen lernen 3 Latein und 12 Französisch.

Wenn du einen Jungen auswählst, ist es viel wahrscheinlicher, dass er Latein lernt, als wenn du ein Mädchen auswählst. Denn bei den Jungen lernen 8 von 10 Latein (also 80 %) und bei den Mädchen nur 3 von 15 (das sind nur 20 %).



Wie wahrscheinlich es ist, dass ein Schüler Latein lernt, hängt also davon ab, ob du einen Jungen oder ein Mädchen auswählst. Die beiden Ereignisse sind also abhängig.

Besteht kein Zusammenhang, nennst du sie unabhängig . Das ist zum Beispiel bei zweimal Würfeln der Fall.

Aufgabe:

Berechnen Sie die Wahrscheinlichkeiten für:

- Klasse - Jungen die Latein lernen
- Klasse - Mädchen die Latein lernen
- Klasse – Schüler die Latein lernen

Schreiben Sie ein Programm in Python um diese Berechnungen zu bestätigen. Denken Sie daran, wie sich die Anzahl der Durchläufe auf die Genauigkeit des Ergebnisses auswirkt.

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

```
Durchläufe  :   100000
            Zähler      % Gesamt        % 2.Ebene
=============================================================
Jungen     : 39944        39.94 %
=============================================================
Latein J   : 31971        31.97 %         80.04 %
Franz. J   : 7973          7.97 %         19.96 %
=============================================================
Mädchen    : 60057        60.06 %
=============================================================
Latein M   : 11898        11.90 %         19.81 %
Franz. M   : 48159        48.16 %         80.19 %
=============================================================
Latein ges: 43869         43.87 %
Franz. ges: 56132         56.13 %
```

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

2.6 Binomialkoeffizient

Stell dir vor, du möchtest deine Wand in zwei Farben streichen. Zur Auswahl stehen gelb, orange, rot und grau. Du wählst zufällig zwei Farben aus. Es gibt dann viele verschiedene Kombinationsmöglichkeiten, zum Beispiel gelb–orange, orange–rot, rot–grau, ... Aber wie viele sind es genau?

Das verrät dir der Binomialkoeffizient . Er besteht aus zwei Zahlen, die übereinander in einer Klammer stehen. Oben steht die Anzahl der Möglichkeiten insgesamt (Abkürzung: n). Unten steht die Anzahl der Farben, die du auswählst (Abkürzung: k):

$$\text{Kombinationsmöglichkeiten} = \binom{n}{k} = \binom{4}{2}$$

Du kannst den Binomialkoeffizienten ganz leicht mit einer Formel berechnen. Das Ausrufezeichen steht dabei für die Fakultät :

$$\binom{n}{k} = \frac{n!}{k!*(n-k)!} = \frac{4!}{2!*(4-2)!} = \frac{4*3*2*1}{(2*1)*(2*1)} = \frac{24}{4} = 6$$

Es gibt also 6 verschiedene Möglichkeiten, die Farben zu kombinieren.

Tipp: Wahrscheinlich hast du auf deinem Taschenrechner eine Taste für den Binomialkoeffizienten, nämlich **nCr**. Dann musst du ihn nicht per Hand ausrechnen!

Der Binomialkoeffizient spielt eine große Rolle bei der Binomialverteilung und der Bernoulli-Formel . Damit kannst du dann nicht nur die Anzahl der Möglichkeiten ausrechnen, sondern auch Wahrscheinlichkeiten. Besonders bei langen Zufallsexperimenten mit sehr vielen Durchgängen hilft sie dir weiter.

2.7 Weitere Beispiele

Jetzt kennst du schon viele wichtige Begriffe der Wahrscheinlichkeitsrechnung. Schau dir zum Schluss noch zwei typische Beispiele an: Das Urnenmodell und die Lottoziehungen.

2.7.1 Wahrscheinlichkeitsrechnung am Urnenmodell

Zufallsexperimente werden oft abstrahiert, indem sie mit dem blinden Ziehen gleich großer und gleich schwerer, aber verschiedenfarbiger Kugeln aus einer Urne verglichen werden, wobei es zwei verschiedene Klassen dieses Experiments gibt: Ziehen mehrerer Kugeln mit Zurücklegen oder ohne Zurücklegen.

Das mehrfache Werfen eines Würfels entspricht beispielsweise dem Ziehen mit Zurücklegen, da dieselbe Zahl ja beliebig oft gewürfelt werden kann. Auch eine Untersuchung, an welchen Tagen Menschen in einer Gruppe Geburtstag haben, entspricht dieser Variante, da auch mehrere Menschen am selben Tag Geburtstag haben können (eine beliebte Aufgabenquelle der Wahrscheinlichkeitsrechnung).

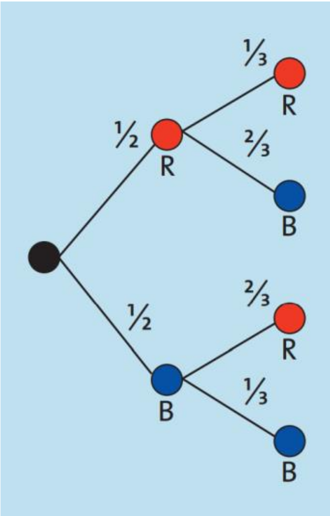
Die Ziehung der Lottozahlen entspricht dagegen dem Ziehen ohne Zurücklegen, da jede der 49 Zahlen nur einmal gezogen werden kann.

Stellen Sie sich im Folgenden eine Urne mit vier Kugeln vor, von denen zwei rot und zwei blau sind.

Beim Ziehen *mit Zurücklegen* wird jeweils eine Kugel gezogen, ihre Farbe wird notiert, und vor der nächsten Ziehung wird sie wieder in die Urne zurückgelegt. Wie groß ist die Wahrscheinlichkeit, bei zwei aufeinanderfolgenden Ziehungen zwei blaue Kugeln zu ziehen? Bei zwei blauen Kugeln von vier Kugeln insgesamt, liegt die Wahrscheinlichkeit für eine einzelne blaue Kugel bei $\frac{2}{4} = \frac{1}{2}$, und da dieses Ereignis zweimal auftreten muss, ist die Wahrscheinlichkeit dafür $\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$.

Beim Ziehen *ohne Zurücklegen* bleibt eine Kugel dagegen dauerhaft draußen, bevor die nächste gezogen wird. Dadurch wird die Menge möglicher Kugeln in jeder Runde um 1 vermindert. Die Wahrscheinlichkeit für eine blaue Kugel in der ersten Runde beträgt wieder $\frac{2}{4} = \frac{1}{2}$. Da sich danach noch eine blaue und zwei rote Kugeln in der Urne befinden, ist die Wahrscheinlichkeit für Blau $\frac{1}{3}$ und für Rot $\frac{2}{3}$. Die Wahrscheinlichkeit für zweimal Blau ist hier also $\frac{1}{2} * \frac{1}{3} = \frac{1}{6}$.

Der entsprechende Entscheidungsbaum sieht wie folgt aus:



Mithilfe des Baumdiagramms lassen sich auf einfache Art und Weise zusammengesetzte Wahrscheinlichkeiten berechnen. Hier sehen Sie einige Beispiele:

- Für die Wahrscheinlichkeit, mindestens eine rote Kugel zu ziehen, gibt es drei mögliche Pfade: Rot-Rot, Rot-Blau und Blau-Rot. Die Wahrscheinlichkeiten dieser Pfade sind für zweimal Rot $\frac{1}{2} * \frac{1}{3} = \frac{1}{6}$, für Rot-Blau $\frac{1}{2} * \frac{2}{3} = \frac{1}{3}$ und noch einmal für Blau-Rot $\frac{1}{2} * \frac{2}{3} = \frac{1}{3}$. Die Summe dieser Wahrscheinlichkeiten ist $\frac{1}{6} + \frac{1}{3} + \frac{1}{3} = \mathbf{\frac{5}{6}}$.
- Die Wahrscheinlichkeit, genau eine blaue Kugel zu ziehen, ergibt sich aus den beiden Pfaden Rot-Blau mit der bereits berechneten Wahrscheinlichkeit $\frac{1}{3}$ und BlauRot mit ebenfalls $\frac{1}{3}$. Die Gesamtwahrscheinlichkeit beträgt also $\mathbf{\frac{2}{3}}$.
- Die Wahrscheinlichkeit für zwei gleiche Kugeln ist $\mathbf{\frac{1}{3}}$, kombiniert aus $\frac{1}{6}$ für zweimal Rot und $\frac{1}{6}$ für zweimal Blau.

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

2.7.2 Wahrscheinlichkeitsrechnung beim Lotto

Beim Lotto wählst du aus 49 Zahlen 6 aus. Wie groß ist die Wahrscheinlichkeit, dass du alle 6 richtig hast? Das kannst du mit dem Binomialkoeffizienten berechnen. Dabei ist die Anzahl der Möglichkeiten insgesamt 49 (n = 49) und du wählst 6 aus (k = 6). Insgesamt gibt es also

$$\binom{49}{6} = 13.983.816$$

Möglichkeiten. Weil nur eine Kombination davon genau richtig ist, liegt die Wahrscheinlichkeit für 6 Richtige bei

$$\frac{1}{13.983.816} = 0{,}000\,000\,072$$

Du kannst dir das Lottospiel auch als Urnenmodell vorstellen. In deiner Urne sind dann 49 Kugeln, beschriftet mit Zahlen. Du ziehst 6 davon.

2.8 Wahrscheinlichkeitsrechnung und Mengenlehre

Diese Beziehungen zwischen Zufallsereignissen lassen sich auch mengentheoretisch im Sinne der Kombinatorik betrachten. Die Menge aller vorstellbaren Ereignisse wird dabei Ω (der griechische Großbuchstabe Omega) genannt. Beim Werfen eines Würfels gilt beispielsweise:

$\Omega = \{1,2,3,4,5,6\}$

Ein erwünschtes Ereignis A ist eine Teilmenge von Ω, also $A \subseteq \Omega$. Beispielsweise ist das Ereignis »eine Sechs wird gewürfelt« die Menge {6}, während das Ereignis »ungerade Zahl« der Menge {1, 3, 5} entspricht.

Mehrere erwünschte Ereignisse, deren Wahrscheinlichkeiten addiert werden, lassen sich als Vereinigungsmengen darstellen. Soll beispielsweise eine 1 oder eine gerade Zahl gewürfelt werden, ist das Gesamtereignis aus A = {1} und B = {2,4,6} die Vereinigungsmenge $A \cup B = \{1\} \cup \{2,4,6\} = \{1,2,4,6\}$.

Entsprechend kann ein Gesamtereignis, das mehrere verschiedene Eigenschaften von Teilereignissen betrachtet, als Schnittmenge geschrieben werden, deren Wahrscheinlichkeit durch Multiplikation der ursprünglichen Wahrscheinlichkeiten berechnet wird. Das Ereignis, aus einem Standardkartenspiel mit 32 Karten ein Pik-As zu ziehen, kann beispielsweise wie folgt definiert werden:

A = { ♠A, ♠K, ♠D, ♠B, ♠10, ♠9, ♠8, ♠7} alle Pik

B = { ♣A, ♠A, ♥A, ♦A} alle Asse

$A \cap B$ = { ♠A}

Es gibt auch weitere Wege zum Ziel. (welche fallen Ihnen ein?)

Weiterhin können Reihen von Zufallsexperimenten als kartesische Produkte betrachtet werden. Das zweimalige Werfen einer Münze ist beispielsweise das kartesische Produkt aus den beiden Mengen M1 und M2 mit den Elementen {K,Z} für Kopf beziehungsweise Zahl:

$\Omega = M_1 \times M_2 = \{K,Z\} \times \{K,Z\} = \{(K,K),(K,Z),(Z,K),(Z,Z)\}$

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

Das Gegenereignis Ā zu einem Ereignis A ist schließlich die Differenzmenge aus Ω und der nicht erwünschten Teilmenge. Soll bei einem zweimaligen Münzwurf etwa mindestens einmal Zahl geworfen werden, so ist dies gleichbedeutend mit:

$\Omega \setminus \{(K,K)\} = \{(K,Z),(Z,K),(Z,Z)\}$

2.9 Satz von der totalen Wahrscheinlichkeit

Mit dem **Satz der totalen Wahrscheinlichkeit** lässt sich die Wahrscheinlichkeit des Ereignisses A berechnen, wenn man nur die **bedingte oder gemeinsame Wahrscheinlichkeit** abhängig von einem zweiten Ereignis B gegeben hat. Manchmal ist auch vom so genannten **Gesetz der totalen Wahrscheinlichkeit** die Rede.

2.9.1 Satz der totalen Wahrscheinlichkeit Formel

Es geht also darum, die gesamte Wahrscheinlichkeit eines Ereignisses A zu berechnen. Mathematisch wird das Gesetz der totalen Wahrscheinlichkeit meistens so aufgeschrieben:

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(B) * P(A|B) + P(\bar{B}) * P(A|\bar{B})$$

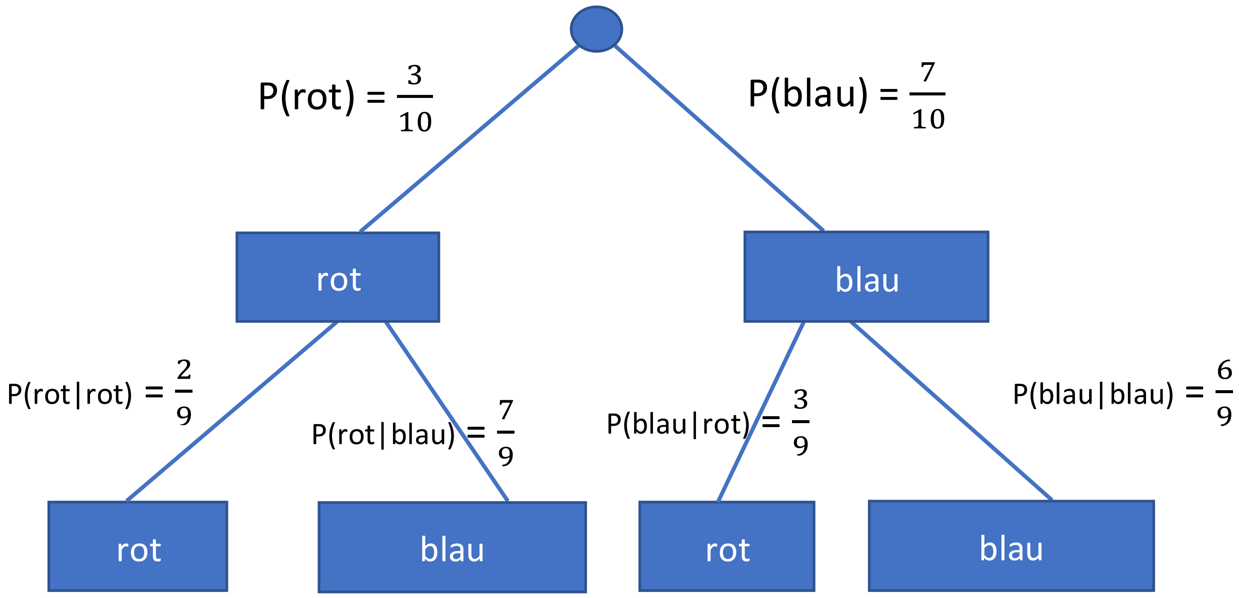
2.9.2 Beziehung zum Satz von Bayes

Außerdem begegnet in der Stochastik einem in der Verbindung mit dem Satz der totalen Wahrscheinlichkeit oft der so genannte **Satz von Bayes** . Mithilfe der Bayes Formel kann die **bedingte Wahrscheinlichkeit** zweier Ereignisse bestimmt werden, welche im Anschluss zur Berechnung der totalen Wahrscheinlichkeit herangezogen wird.

2.9.3 Satz der totalen Wahrscheinlichkeit Beispiel

Anhand eines Beispiels wird das ganze gleich viel verständlicher. Stell dir vor, in einer Urne befinden sich 3 rote und 7 blaue Kugeln. Jetzt ziehst du zwei Kugeln ohne Zurücklegen aus der Urne heraus. Wie hoch ist die Wahrscheinlichkeit, dass genau eine der beiden Kugeln blau ist?

Baumdiagramm:



Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

2.9.4 Wahrscheinlichkeit berechnen

Nun können wir mit Hilfe des Baumdiagramms ganz einfach die Werte in unsere Formel eintragen und erhalten:

P(1 x blau) = P(blau) * P(blau|rot) + P(rot) * P(rot|blau) =

$$\frac{7}{10} * \frac{3}{9} + \frac{3}{10} * \frac{7}{9} = \frac{42}{90} = 46{,}6\ \%$$

2.10 Bayes Theorem

Der **Satz von Bayes** gehört zu den wichtigsten Sätzen der Wahrscheinlichkeitsrechnung. Er ermöglicht es die bedingte Wahrscheinlichkeit zweier Ereignisse A und B zu bestimmen, falls eine der beiden bedingten Wahrscheinlichkeiten bereits bekannt ist. Dieser mathematische Satz ist auch unter den Namen **Formel von Bayes** oder **Bayes Theorem** bekannt.

2.10.1 Satz von Bayes Formel

Die mathematische Formel für den Satz von Bayes sieht so aus:

$$\mathrm{P(A|B)} = \frac{P(B|A)*P(A)}{P(B)}$$

Hier ist P(A|B) die bedingte Wahrscheinlichkeit von A falls B bereits eingetreten ist. Analog steht P(B|A) für die Wahrscheinlichkeit des Ereignisses von B unter der Bedingung, dass A eingetreten ist. P(A) und P(B) stehen jeweils für die jeweiligen Wahrscheinlichkeiten der Ereignisse.

Satz von Bayes einfach erklärt

Wenn man also die Wahrscheinlichkeit von B unter der Bedingung von A gegeben hat kann man mit der Bayes Formel auch die bedingte Wahrscheinlichkeit berechen, dass A eintritt, wenn B bereits eingetreten ist. Einfach gesagt ermöglicht der Satz von Bayes es Schlussfolgerungen von der anderen Seite aus zu betrachten: Man geht von dem bekannten Wert P(B|A) aus, ist aber eigentlich an dem Wert P(A|B) interessiert. Der Satz von Bayes folglich berechnet die umgekehrte Form der gegebenen bedingten Wahrscheinlichkeit.

2.10.2 Satz von Bayes Beispiel

Schauen wir uns am besten gleich ein praktisches Beispiel dazu an. Stell dir vor, ein Kommilitone von dir wird nach dem Feiern von der Polizei aufgehalten und muss einen Alkoholtest machen. Bei Personen, die tatsächlich Alkohol getrunken haben, erkennt der Test das in 99,9% der Fälle.

Allerdings liefert er auch in 3% der Fälle ein positives Ergebnis, obwohl die getestete Person keinen Alkohol getrunken hat. Wir wissen also:

P(+|Alkohol)= 0,999

P(+|kein Alkohol)=0,03

Außerdem wissen wir, dass 5% der getesteten Personen tatsächlich Alkohol konsumiert haben:

P(Alkohol)=0,05

Die Wahrscheinlichkeit, dass eine getestete Person keinen Alkohol getrunken hat, liegt also bei 95%. (100% - 5%)

Der Test fällt bei deinem Kommilitonen positiv aus. Wie hoch ist die Wahrscheinlichkeit, dass er tatsächlich Alkohol konsumiert hat?

Diese Frage lässt sich mit Hilfe des Satzes von Bayes beantworten. Die beiden Wahrscheinlichkeiten, die wir im Zähler der Formel einsetzen müssen, haben wir gegeben. Allerdings fehlt uns noch die Wahrscheinlichkeit dafür, dass der Test positiv ausfällt. Da wir aber die beiden bedingten Wahrscheinlichkeiten gegeben haben, können wir das mit Hilfe des Satzes der totalen Wahrscheinlichkeit herleiten. Ein positives beziehungsweise negatives Testergebnis kürzen wir im Folgenden mit einem Plus beziehungsweise einem Minus ab.

2.10.3 Satz von Bayes Anwendung

So, jetzt müssen wir nur noch alle Werte in die Formel von vorhin einsetzen.

P(+|Alkohol)= 0,999 pos. Test bei Alkohol Konsum

P(+|kein Alkohol)=0,03 pos. Test aber kein Alkohol Konsum

P(Alkohol)=0,05 Alkohol Konsum bei get. Personen

P(kein Alkohol)=0,95 kein Alkohol Konsum bei get. Personen

Wahrscheinlichkeit dafür, dass der Test positiv ausfällt:

P(+) = P(Alkohol) * P(+|Alkohol) + P(kein Alkohol) * P(+|kein Alkohol)

= 0,05 * 0,999 + 0,95 *0,03

= 0,07845

$$P(\text{Alkohol}|+) = \frac{P(+|Alkohol) * P(Alkohol)}{P(+)} = \frac{0{,}999 * 0{,}05}{0{,}07845} = 0{,}6371$$

Da der Test positiv ausgefallen ist, hat dein Kommilitone also mit einer Wahrscheinlichkeit von 63,67% tatsächlich Alkohol getrunken.

2.11 Vierfeldertafel

Die Vierfeldertafel ist ein Hilfsmittel in der Stochastik, um Zusammenhänge zwischen zwei Ereignissen darzustellen.

An ihr kann man neue Informationen (zum Beispiel Wahrscheinlichkeiten, oder absolute Häufigkeiten) ablesen.

Die Vierfeldertafel hilft auch, die Unabhängigkeit von Ereignissen zu untersuchen.

2.11.1 Allgemeine Form

Im Folgenden werden die Einträge der Vierfeldertafel erklärt. Die Buchstaben A und B bezeichnen dabei zwei Ereignisse (man hätte aber genauso gut andere Buchstaben nehmen können). $\bar{A}$ und $\bar{B}$ stehen für ihre Gegenereignisse.

2.11.2 Beschriftung

Egal, ob man Wahrscheinlichkeiten oder absolute Häufigkeiten beschreiben will,

- stehen in der ersten Zeile die Symbole für das Ereignis A und sein Gegenereignis $\bar{A}$.
- in der ersten Spalte die Symbole für das Ereignis B und sein Gegenereignis $\bar{B}$.

	$\boldsymbol{A}$	$\boldsymbol{\bar{A}}$	
$\boldsymbol{B}$			
$\boldsymbol{\bar{B}}$			

2.11.2.1 Vierfeldertafel mit Wahrscheinlichkeiten

- In der letzten Zeile der A-Spalte steht die Wahrscheinlichkeit für das Ereignis A. Man bezeichnet diese mit $P(A)$
- In der letzten Zeile der $\bar{A}$ -Spalte steht die Wahrscheinlichkeit für das Ereignis $\bar{A}$. Diese wird mit $P(\overline{A})$ bezeichnet.

	A	$\bar{A}$	
B			
$\bar{B}$			
	$\boldsymbol{P(A)}$	$\boldsymbol{P(\bar{A})}$	

- In der letzten Spalte der B -Zeile steht die Wahrscheinlichkeit $P(B)$ für das Ereignis B.
- In der letzten Spalte der $\overline{\mathrm{B}}$-Zeile steht die Wahrscheinlichkeit $\mathrm{P}(\overline{\mathrm{B}})$ für das Ereignis $\overline{\mathrm{B}}$.

	A	$\bar{A}$	
B			$\boldsymbol{P(B)}$
$\bar{B}$			$\boldsymbol{P(\bar{B})}$
	$P(A)$	$P(\bar{A})$	

- Da die Ereignisse A und $\bar{A}$ Gegenereignisse sind, müssen sich ihre Wahrscheinlichkeiten zu 1 addieren.
- Dasselbe gilt für B und $\bar{B}$.

Daher ist der Eintrag rechts unten in der Vierfeldertafel immer eine 1.

	A	$\bar{A}$	
B			$P(B)$
$\bar{B}$			$P(\bar{B})$
	$P(A)$	$P(\bar{A})$	**1**

2.11.2.2 Innere vier Felder

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

- In den restlichen vier Feldern notiert man die Wahrscheinlichkeiten der Schnittmengen der Ereignisse $A \cap B$, $\bar{A} \cap$ B, $A \cap \bar{B}$, $\bar{A} \cap \bar{B}$.
- Dabei ist die letzte Wahrscheinlichkeit in einer Zeile die Summe der anderen beiden. Zum Beispiel $P(B)$ = $P(A \cap B)$ + $P(\bar{A} \cap B)$.
- Dasselbe gilt für die letzte Wahrscheinlichkeit einer Spalte. Zum Beispiel $P(\bar{A})$ = $P(\bar{A} \cap B)$ + $P(\bar{A} \cap \bar{B})$.

	A	$\bar{A}$	
B	$\boldsymbol{P(A \cap B)}$	$\boldsymbol{P(\bar{A} \cap B)}$	$P(B)$
$\bar{B}$	$\boldsymbol{P(A \cap \bar{B})}$	$\boldsymbol{P(\bar{A} \cap \bar{B})}$	$P(\bar{B})$
	$P(A)$	$P(\bar{A})$	1

2.11.3 Eigenschaften und Rechenregeln

2.11.3.1 Vierfeldertafel mit Wahrscheinlichkeiten

Wie man oben bereits gesehen hat, gilt:

- Jede Wahrscheinlichkeit in der untersten Zeile ist die Summe der beiden Wahrscheinlichkeiten darüber.
- Jede Wahrscheinlichkeit in der letzten Spalte ist die Summe der beiden Wahrscheinlichkeiten links davon.
- Die letzte Zeile und die letzte Spalte müssen jeweils in der Summe 1 ergeben.

2.11.4 Übungen: Fülle mit den folgenden Informationen eine Vierfeldertafel aus!

- $P(A)$=0,45 ; $P(A \cap \bar{B})$=0,2 ; $P(\bar{B})$=0,7

	A	$\bar{A}$	
B	$\boldsymbol{P(A \cap B)}$	$\boldsymbol{P(\bar{A} \cap B)}$	$P(B)$
$\bar{B}$	$\boldsymbol{P(A \cap \bar{B})}$	$\boldsymbol{P(\bar{A} \cap \bar{B})}$	$P(\bar{B})$
	$P(A)$	$P(\bar{A})$	1

	A	$\bar{A}$	
B			
$\bar{B}$			
			1

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

- $P(A \cap B)$=0,12 ; $P(\bar{A})$=0,51 ; $P(B)$=0,44

➢	A	$\bar{A}$	
B	$\boldsymbol{P(A \cap B)}$	$\boldsymbol{P(\bar{A} \cap B)}$	$P(B)$
$\bar{B}$	$\boldsymbol{P(A \cap \bar{B})}$	$\boldsymbol{P(\bar{A} \cap \bar{B})}$	$P(\bar{B})$
	$P(A)$	$P(\bar{A})$	1

	A	$\bar{A}$	
B			
$\bar{B}$			
			1

- $P(B)$=0,99 ; $P(A \cap \bar{B})$=0,002 ; $P(\bar{A} \cap B)$=0,85

➢	A	$\bar{A}$	
B	$\boldsymbol{P(A \cap B)}$	$\boldsymbol{P(\bar{A} \cap B)}$	$P(B)$
$\bar{B}$	$\boldsymbol{P(A \cap \bar{B})}$	$\boldsymbol{P(\bar{A} \cap \bar{B})}$	$P(\bar{B})$
	$P(A)$	$P(\bar{A})$	1

	A	$\bar{A}$	
B			
$\bar{B}$			
			1

2.12 Absolute Häufigkeit

Die absolute Häufigkeit gibt an, wie oft ein bestimmtes Ereignis bei mehrmaliger Wiederholung eines Zufallsexperiments eintritt. Als Anzahl ist sie immer eine natürliche Zahl zwischen 0 und der Gesamtzahl von Versuchen.

Bei Experimenten mit zwei Ereignissen kann man die absolute Häufigkeit in einer Vierfeldertafel darstellen.

Beispiel:

Wenn ein Würfel 20-mal geworfen wird und fünfmal eine 3 fällt, so ist die absolute Häufigkeit von 3 gleich 5.

Für die absolute Häufigkeit wird oft der Buchstabe H verwendet. Für das Beispiel mit dem Würfel schreibt man also H(3)=5.

2.12.1 Darstellung von absoluten Häufigkeiten

Absolute Häufigkeiten lassen sich je nach Anwendung gut in Tabellen festhalten. Typisch sind Strichlisten, in denen pro eingetretenen Ereignis ein Strich geschrieben wird. Wenn du die auftretenden Augenzahlen beim Wurf eines Würfels aufschreibst, könnte deine Strichliste beispielsweise so aussehen:

<table>
<tr><th>Augenzahl</th><th>absolute Häufigkeit</th></tr>
<tr><td>1</td><td>||||</td></tr>
<tr><td>2</td><td>||</td></tr>
<tr><td>3</td><td>𝍸||</td></tr>
<tr><td>4</td><td>𝍸</td></tr>
<tr><td>5</td><td>|||</td></tr>
<tr><td>6</td><td>|</td></tr>
</table>

Absolute Häufigkeiten lassen sich aber auch in Vierfeldertafeln darstellen. Durch diese Darstellung lassen sich Zusammenhänge zwischen den Ereignissen abbilden. Du kannst so zum Beispiel festhalten, wie viele Kinder und wie viel Erwachsene bei einer Umfrage angeben, Spinat zu mögen und wie viele keinen Spinat mögen:

	mag Spinat	mag keinen Spinat	
Kind	8	14	22
Erwachsener	47	31	78
	55	45	100

2.13 Beziehung zu relativer Häufigkeit

Im Gegensatz zur absoluten Häufigkeit berücksichtigt die relative Häufigkeit auch die **Anzahl der Versuche**.

Zwischen der absoluten Häufigkeit und der relativen Häufigkeit gilt folgende Beziehung:

absolute Häufigkeit=relative Häufigkeit * Anzahl der Versuche

$$absolute\ Häufigkeit = relative\ Häufigkeit * Anzahl\ der\ Versuche$$

$$\rightarrow \qquad relative\ Häufigkeit = \frac{absolute\ Häufigkeit}{Anzahl\ der\ Versuche}$$

In dem oberen Beispiel mit dem Würfel heißt das, $\frac{5}{20} = 0{,}25 = 25\ \%$

2.13.1.1 Vierfeldertafel mit absoluten Häufigkeiten

2.13.1.2 Äußere Felder

- In der letzten Zeile der A-Spalte steht die absolute Häufigkeit für das Ereignis A. Man bezeichnet diese mit $H(A)$.
- In der letzten Zeile der $\bar{A}$-Spalte steht die absolute Häufigkeit für das Ereignis $\bar{A}$. Man bezeichnet diese mit $H(\bar{A})$.

	A	$\bar{A}$	
B			
$\bar{B}$			
	$\boldsymbol{H(A)}$	$\boldsymbol{H(\bar{A})}$	

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

- In der letzten Spalte der B-Zeile steht die absolute Häufigkeit $H(B)$ für das Ereignis B.
- In der letzten Spalte der $\bar{B}$-Zeile steht die absolute Häufigkeit $H(\bar{B})$ für das Ereignis $\bar{B}$.

	A	$\bar{A}$	
B			$\boldsymbol{H(B)}$
$\bar{B}$			$\boldsymbol{H(\bar{B})}$
	$H(A)$	$H(\bar{A})$	

Der Eintrag unten rechts einer Vierfeldertafel mit absoluten Häufigkeiten ist die Gesamtzahl G an Versuchen, oder an untersuchten Objekten.

- die absoluten Häufigkeiten von A und $\bar{A}$ zu G addieren, sowie auch
- die absoluten Häufigkeiten von B und $\bar{B}$.

	A	$\bar{A}$	
B			$H(B)$
$\bar{B}$			$H(\bar{B})$
	$H(A)$	$H(\bar{A})$	$\boldsymbol{G}$

2.13.1.3 Innere vier Felder

- In diese vier Felder notiert man die absoluten Häufigkeiten der Schnittmengen der Ereignisse $A \cap B$, $\bar{A} \cap$ B, $A \cap \bar{B}$, $\bar{A} \cap \bar{B}$.
- Dabei ist die letzte absolute Häufigkeit in einer Zeile die Summe der anderen beiden. Zum Beispiel $H(B) = H(A \cap B) + H(\bar{A} \cap B)$.
- Dasselbe gilt für die letzte absolute Häufigkeit einer Spalte. Zum Beispiel $H(\bar{A}) = H(\bar{A} \cap B) + H(\bar{A} \cap \bar{B})$.

	A	$\bar{A}$	
B	$\boldsymbol{H(A \cap B)}$	$\boldsymbol{H(\bar{A} \cap B)}$	$H(B)$
$\bar{B}$	$\boldsymbol{H(A \cap \bar{B})}$	$\boldsymbol{H(\bar{A} \cap \bar{B})}$	$H(\bar{B})$
	$H(A)$	$H(\bar{A})$	G

2.13.2 Eigenschaften und Rechenregeln

2.13.2.1 Vierfeldertafel mit absoluten Häufigkeiten

- Jede absolute Häufigkeit in der untersten Zeile ist die Summe der beiden absoluten Häufigkeiten darüber.
- Jede absolute Häufigkeit in der letzten Spalte ist die Summe der beiden absoluten Häufigkeiten links davon.
- Die letzte Zeile und die letzte Spalte müssen jeweils in der Summe G ergeben.

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

Bemerkung: Statt Wahrscheinlichkeiten und absoluten Häufigkeiten, kann man auch die Prozentwerte der jeweiligen Ereignissen in eine Vierfeldertafel eintragen. In solchen Fällen kommt nicht mehr 1 oder G als Summe heraus, sondern 100%.

2.13.3 Übungen: Fülle mit den folgenden Informationen eine Vierfeldertafel mit absoluten Häufigkeiten aus!

- In einer Schulklasse mit 29 Schüler*innen haben 10 Schüler*innen braune Haare und 7 Schüler*innen grüne Augen. 5 Schüler*innen haben grüne Augen und braune Haare.

A: grüne Augen

B: Braune Haare

	A	$\bar{A}$	
B			
$\bar{B}$			

- Am Sportunterricht nehmen insgesamt 25 Kinder teil, wovon 15 weiblich sind. Genau 15 Kinder sind gut im Weitwurf. 10 Mädchen sind gut im Weitwurf.

A: Weiblichkeit

B: Weitwurf

	A	$\bar{A}$	
B			
$\bar{B}$			

- Bei einer Versuchsreihe nehmen 47 Personen teil. 20 von diesen Personen wurden auf eine bestimmte Krankheit positiv getestet. 32 Testpersonen sind gegen diese Krankheit geimpft, wobei 15 Personen positiv getestet wurden und nicht dagegen geimpft sind.

A: positiv

B: geeimpft

	A	$\bar{\mathrm{A}}$	
B			
$\bar{\mathrm{B}}$			

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

- In einer Firma arbeiten 45 Männer und 50 Frauen. 30 weibliche Mitarbeitende der Firma sind jünger als 50 Jahre und 27 Männer sind älter als 50 Jahre.

A: Männer

B: jünger

	A	$\bar{A}$	
B			
$\bar{B}$			

2.14 Normalverteilung

Die Normalverteilung oder auch Gauß-Verteilung bzw. Gaußsche Normalverteilung ist ein elementarer Begriff aus der Stochastik. Sie ist die wichtigste Wahrscheinlichkeitsverteilung und wird zur Analyse von Daten, aber auch zur Berechnung von Wahrscheinlichkeiten hergenommen.

Wahrscheinlichkeitsverteilungen werden immer hergenommen, wenn man ein Zufallsexperiment modellieren, auswerten oder berechnen möchte.

Anhand der Verteilung können die Wahrscheinlichkeiten für einzelne Ausprägungen einer Zufallsvariable berechnet werden, aber auch bei Bekanntheit der Verteilung einzelne Parameter abgeleitet werden. Die Normalverteilung dient auch als Basis für weitere statistische Auswertungen wie beispielsweise t-Tests oder Regressions- und Varianzanalysen.

Die Normalverteilung wird meist für sehr große Stichproben hergenommen oder dient als Hilfsmittel, wenn die eigentliche Verteilung unbekannt ist.

Anwendung findet sie in vielen verschiedenen Fällen in den Natur-, Geistes- oder Wirtschaftswissenschaften:

- Biologie (Körpergröße, Länge von Gliedmaßen, Blutdruck)
- Meteorologie (Regenmenge, Sonnenstunden)
- Physik (Messfehler)
- Produktion (Kapazitäten, Qualitätssicherung, Füllmenge von Packungen)
- Gesellschaftswissenschaften (Einkommen (bei logarithmierten Werten), Intelligenz, Sozialkompetenz)
- Finanzmarkt (Preisänderung von Aktien)

2.14.1 Die Funktion der Normalverteilung

Die Normalverteilung kann einerseits als Verteilung, als Verteilungsfunktion oder als grafische Darstellung vorkommen. Die Normalverteilung hat eine sehr charakteristische Form, weshalb sie auch Glockenkurve oder Gaußsche Glockenkurve genannt wird.

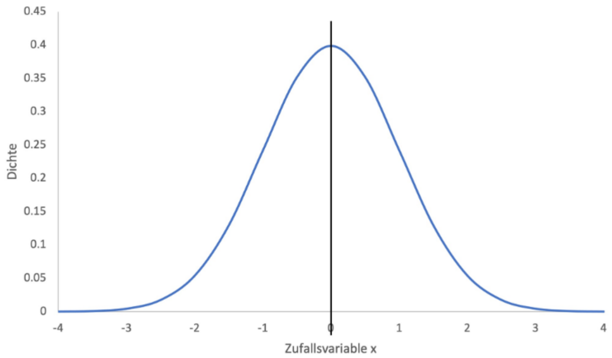
Wichtige Bestandteile der Formel sind die Variable x, der Erwartungswert μ und die Varianz σ^2 bzw. die Standardabweichung σ. Die Normalverteilung errechnet sich anhand folgender Formel:

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

$$f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

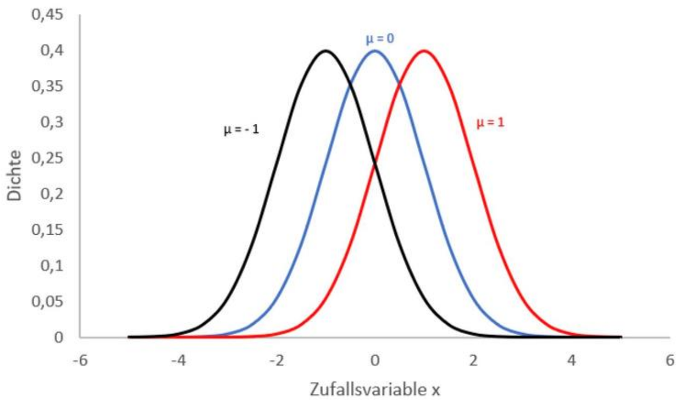
Parameter	Definition	Erklärung	Grafische Auswirkung
x	**Zufallsvariable**	-	Werte auf der X-Achse
µ	**Erwartungswert (mu)**	Die Ausprägung, die mit der häufigsten Wahrscheinlichkeit auftritt. Im Fall der Normalverteilung identisch mit dem Median.	Lage und Maximum des Graphen - µ<0; Graph verschiebt sich nach links - µ>0; Graph verschiebt sich nach rechts
σ^2	**Varianz (sigma²)**	Maß für die Streuung	Form des Graphen: - σ > 1 Stauchung (breiter) - 0 < σ < 1 Streckung (schmaler)
σ	**Standardabweichung - Wurzel aus der Varianz σ²**	Gibt die absolute Streuung an.	

Sind Erwartungswert µ = 0 und Standardabweichung σ =1, so spricht man von der sogenannten Standardnormalverteilung. Diese wird auch oft mit dem griechischen Buchstaben φ (phi) angegeben.



Sind die Werte µ und σ ungleich 0 ergeben sich quasi unendlich viele Varianten der Normalverteilung.

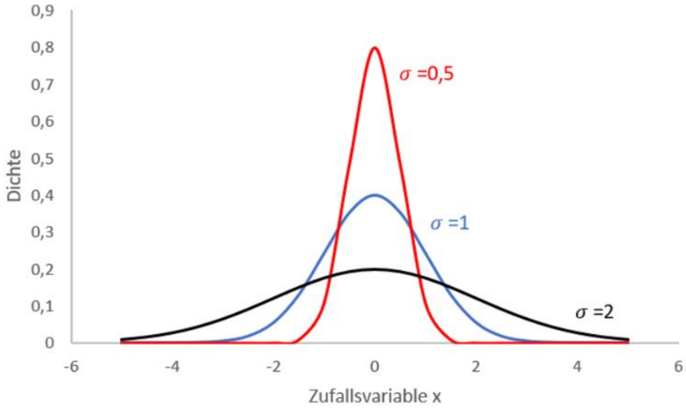
Im Folgenden zeigen wir zwei weitere Ausprägungen von µ und deren Auswirkung auf den Graphen.



Die blaue Kurve zeigt die Standardnormalverteilung mit µ = 0 und σ =1. Die rote Kurve zeigt einen Mittelwert von µ =1 bei σ = 1 und die graue Kurve einen Wert von µ = -1 bei σ =1.

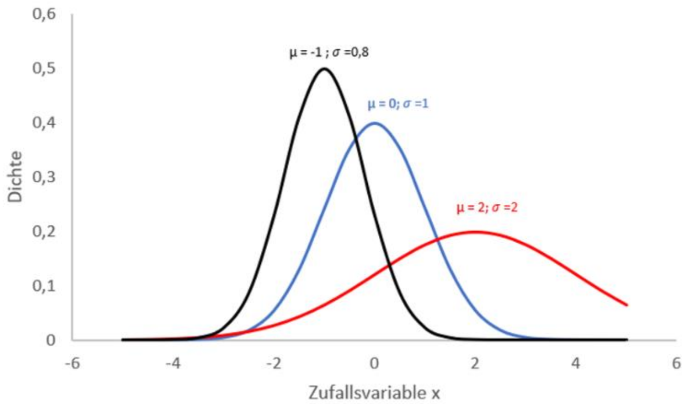
Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

Im Folgenden veranschaulichen wir unterschiedliche Standardabweichungen:



Die blaue Kurve zeigt wieder die Standardnormalverteilung mit $\mu = 0$ und $\sigma = 1$. Die rote Kurve zeigt eine Standardabweichung von $\sigma = 0{,}5$ und die graue Kurve einen Wert $\sigma = 2$, beide bei einem Erwartungswert von $\mu = 0$.

Nun kombinieren wir beide Parameter miteinander:



Die blaue Kurve zeigt wieder eine Standardnormalverteilung mit $\mu = 0$ und $\sigma = 1$. Die rote Kurve zeigt eine Verschiebung von $\mu = 0$ auf $\mu = 2$ und eine Stauchung von $\sigma = 1$ auf $\sigma = 2$. Der Mittelwert liegt somit höher, gleichzeitig auch die Streuung. Der schwarze Graph wiederum zeigt eine Verschiebung von $\mu = 0$ auf $\mu = -1$ und eine Streckung von $\sigma = 1$ auf $\sigma = 0{,}8$, also eine geringere Streuung. Auf diese Weise ergeben sich unendlich viele mögliche Ausprägungen der Normalverteilung.

Weitere Besonderheiten der Normalverteilung

- Die Normalverteilung ist stets symmetrisch mit einer *Symmetrieachse* bei $x = \mu$.
- Die Fläche bzw. das Integral unter der Kurve von $-\infty$ bis $+\infty$ ist stets 1. Dies bedeutet, dass die Wahrscheinlichkeit aller Ausprägungen dieser Verteilung in Summe 1, also 100%, ist.
- Die Normalverteilung ist eine stetige Wahrscheinlichkeitsverteilung, das heißt, dass *theoretisch* alle reellen Zahlen als Ausprägung von x möglich sind.

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

2.14.2 Der zentrale Grenzwertsatz

Durch den zentralen Grenzwertsatz lässt sich erklären, warum die Normalverteilung so vielfältig anwendbar ist.

Grenzwertsatz einfach erklärt

Der Grenzwertsatz sagt aus, dass wenn man nur eine ausreichend große Anzahl an Zufallsvariablen aus derselben Verteilung (unabhängig von der Art der Verteilung) nimmt, diese immer annähernd normalverteilt sein werden.

Der Mittelwert der Stichprobe wird dann immer näherungsweise dem Mittelwert der Grundgesamtheit entsprechen, genau wie die Varianz.

2.14.2.1 Die 68 – 95 – 99,7 Regel

Die 69-95-99,7 Regel gibt Aufschluss darüber, wie viel Prozent der Ausprägungen im Rahmen von ein, zwei oder drei Standardabweichungen liegen.

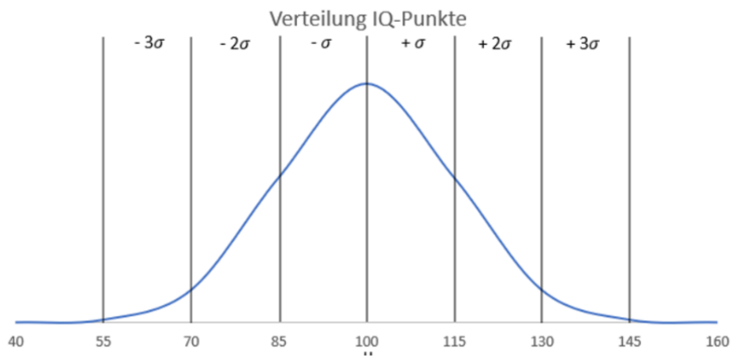
Konkret bedeutet dies, dass bei einer Normalverteilung 68% aller Ausprägungen im Rahmen einer Standardabweichung nach oben nach unten um den Erwartungswert liegen. 95% aller Werte streuen im Rahmen von zwei Standardabweichungen um den Erwartungswert und 99,7% aller Werte liegen um je drei Standardabweichungen um den Mittelwert.

2.14.3 Beispiel für eine Normalverteilung

Betrachten wir die Normalverteilung und ihre Interpretation anhand eines fiktiven, stark vereinfachten Beispiels. 1.000 Erstklässler einer größeren Stadt nehmen an einem IQ-Test teil.

Intelligenz – sofern die Stichprobe unter denselben Rahmenbedingungen zufällig ausgewählt wurde – ist normalverteilt und IQ-Tests sind ab dem Schulalter sinnvoll einsetzbar. Mit 1.000 Kindern liegt auch eine ausreichend große Stichprobengröße vor.

In folgendem Diagramm sind die Werte des IQ-Tests veranschaulicht. Die x-Achse zeigt die IQ-Werte der Kinder und die Y-Achse, mit welcher Wahrscheinlichkeit diese aufgetreten sind. Zur Vereinfachung haben wir die Werte in 15-er Schritten eingeteilt.



Interpretieren wir das Diagramm mit dem bereits gesammelten Wissen, ergeben sich folgende Erkenntnisse:

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

Parameter	
Erwartungswert µ	Die Kinder haben einen durchschnittlichen IQ von 100
Standardabweichung σ	Die Standardabweichung beträgt 15 Punkte
68 – 95 – 99,7 Regel	
+/- 1 σ (68)	68,2% der Kinder haben einen IQ von 85 bis 115, davon 34,1% von 85 bis 100 ($\mu - \sigma$) und 34,1% zwischen 100 und 115 ($\mu + \sigma$)
+/- 2 σ (95)	95% der Kinder haben einen IQ von 70 bis 130, davon 47,7% zwischen 70 und 100 ($\mu - 2\sigma$) und 47,7% zwischen 100 und 130 ($\mu + 2\sigma$)
+/- 3 σ (99,7)	99,7% der Kinder haben einen IQ zwischen 55 und 145, 49,85% davon zwischen 55 und 100 ($\mu - 3\sigma$)und 49,85% zwischen 100 und 145 ($\mu + 3\sigma$)
Zentraler Grenzwert	
Hätten wir nur zehn Kinder getestet, wäre der IQ vermutlich bei einem geringeren oder höheren Wert als 100, beispielsweise bei 90 oder 110, bei fünfzig Kindern nähern wir uns der 100 schon mehr an. Je größer die Stichprobe wird, desto genauer die Annäherung mit der Normalverteilung.	

2.14.4 Fazit

Die Normalverteilung ist die wichtigste und grundlegende Wahrscheinlichkeitsverteilung.

Beachtet man ein paar Rahmenbedingungen wie den zentralen Grenzwert für eine ausreichend große Stichprobengröße und kennt man die Interpretation der verschiedenen Parameter, kann man viele komplexe statistische Probleme mit einer einfachen, allgemeingültigen Verteilung modellieren.

- Permutation ohne Wiederholung

3 Mathematische Statistik

Die in den vorigen Abschnitten besprochenen Zufallsereignisse, aber auch andere Ergebnisse von Zählungen, Messungen, Umfragen und anderen Untersuchungen lassen sich mit den Mitteln der Statistik in mehr oder weniger sinnvolle Zusammenhänge bringen. In diesem Abschnitt werden die wichtigsten Methoden und Funktionen der Statistik eingeführt, von denen auch KI-Algorithmen Gebrauch machen.

3.1 Deskriptive Statistik

Die deskriptive Statistik liefert Werte, mit denen sich Verteilungen beschreiben lassen.

3.1.1 Prozentrechnung

3.1.1.1 Verkehr

Nehmen Sie an, Sie stellen sich einmal zur Hauptverkehrszeit an einem Wochentag und ein anderes Mal an einem Sonntagmorgen je 10 Minuten lang an eine Straße mit mittlerem Verkehrsaufkommen und erstellen die folgende Tabelle der vorbeikommenden Fortbewegungsmittel.

Fortbewegungsmittel	Berufsverkehr	Sonntag
PKW	25	6
LKW	4	1
Busse	2	0
Fahrräder	17	8
Summe	**48**	**15**

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

Verkehrsaufkommen an einer mittelgroßen Straße innerhalb von 10 Minuten, einmal im Berufsverkehr und einmal am Sonntagmorgen

Hieran lässt sich eine erste interessante Methode der Statistik erläutern:

Die Normalisierung von Werten auf vergleichbare Bezugsgrößen. Zunächst wollen Sie nämlich vielleicht einfach wissen, ob sich die Verteilung der Verkehrsmittel wesentlich ändert. Da hilft es wenig, dass im Berufsverkehr insgesamt 48, am Sonntagmorgen jedoch nur 15 Verkehrsmittel vorbeikommen.

Hier hilft die Prozentrechnung weiter:

Wenn die Anzahlen jeweils in Anteile von 100 umgerechnet werden, werden sie miteinander vergleichbar. Dazu dient die Formel Einzelwert ÷ Summe · 100. Der Anteil der PKW im Berufsverkehr beträgt also beispielsweise 25 ÷ 48 · 100 ≈ 52,08 %. Tabelle 2.8 zeigt die Werte aus der vorigen Tabelle in Prozentwerten.

Fortbewegungsmittel	Berufsverkehr	Sonntag
PKW	52%	40%
LKW	8%	7%
Busse	4%	0%
Fahrräder	35%	53%

Die Verkehrsmittel aus der vorigen Tabelle in Prozent des jeweiligen Gesamtverkehrsaufkommens

Sollte die Spaltensumme in einer solchen Tabelle in Einzelfällen nicht genau 100 % betragen, liegt das daran, dass die Werte gerundet werden (hier zum Beispiel auf zwei Nachkommastellen).

Auch die Unterschiede im Gesamtverkehrsaufkommen lassen sich prozentual angeben. Wird der Berufsverkehr als Normalfall mit 100 % betrachtet, dann sind am Sonntagmorgen nur noch 15 ÷ 48 · 100 = 31,25 %, also weniger als ein Drittel an Verkehrsmitteln unterwegs. Umgekehrt beträgt das Spitzenverkehrsaufkommen 320 % (48 ÷ 15 · 100) des Sonntagsverkehrs.

3.1.1.2 Mehrwertsteuer

Prozentrechnung wird auch oft verwendet, um prozentuale Wertänderungen anzugeben. Eines der bekanntesten Beispiele ist die gesetzliche Umsatzsteuer oder Mehrwertsteuer. Diese beträgt in Deutschland für die meisten Waren und Dienstleistungen 19 % und (vereinfacht gesagt) für Lebensnotwendiges 7 %.

Die in Supermärkten, Kaufhäusern und im sonstigen Einzelhandel ausgewiesenen Preise sind Bruttopreise, werden also einschließlich der gesetzlichen Mehrwertsteuer angegeben. Um den Mehrwertsteuerbetrag und den Nettopreis zu berechnen, genügt es nicht, den Mehrwertsteuersatz auf den Bruttopreis anzuwenden, denn er wurde vom Nettopreis berechnet und zu diesem addiert.

Angenommen, ein Buch wie das vorliegende kostet 29,90 €. Für Bücher gilt der ermäßigte Mehrwertsteuersatz von 7 %, also entspricht dieser Preis dem Nettopreis + 7 % dieses Nettopreises. Die 29,90 € sind mit anderen Worten 107 % des Nettopreises, und Sie können diesen und den Mehrwertsteuerbetrag durch folgende Rechnungen ermitteln:

29,90 € ÷ 107 · 100 ≈ 27,94 € ist der Nettopreis, 29,90 € ÷ 107 · 7 ≈ 1,96 € der Mehrwertsteuerbetrag. Bei einem Artikel mit 19 % Mehrwertsteuer, etwa einem Videospiel für 49,90 €, gilt entsprechend: Nettopreis 49,90 € ÷ 119 · 100 ≈ 41,93 €, Mehrwertsteuerbetrag 49,90 € ÷ 119 · 19 ≈ 7,97 €.

3.1.2 Statistikfunktionen

Die wichtigsten Funktionen der Statistik beschäftigen sich mit der Berechnung von Durchschnittswerten und anderen Zusammenfassungen aus Datenreihen, die mit den Oberbegriffen Aggregationen oder Lageparameter bezeichnet werden. Dafür gibt es verschiedene Verfahren, die je nach Datengrundlage unterschiedlich sinnvoll sind. Dieses wichtige Teilgebiet der Statistik wird als deskriptive oder beschreibende Statistik bezeichnet.

3.1.2.1 Minimum, Maximum, Spannweite

Anschauliche Größen für eine Zahlenmenge X sind das Maximum x_{max} und das Minimum x_{min}, also der höchste beziehungsweise niedrigste Wert aus der Gruppe.

Beispiel:

Nehmen Sie beispielsweise an, bei einem Test erreichen die Teilnehmenden folgende Punktzahlen von insgesamt 100 Punkten:

[98, 91, 87, 83, 81, 76, 74, 68, 59]

Dafür gelten folgende Werte:

- $x_{max} = 98$
- $x_{min} = 59$
- $R = x_{max} - x_{min} = 39$ (Spannweite)

3.1.2.2 Arithmetisches Mittel, Mittelwert

Eine weitere recht bekannte Funktion ist das arithmetische Mittel, auch einfach Mittelwert genannt. Dabei werden alle Werte zusammenaddiert, anschließend wird das Ergebnis durch die Anzahl der Werte dividiert:

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

Die durchschnittlich erzielte Punktzahl aus den Testergebnissen ist:

$$\bar{x}_{arithm} = \frac{1}{9} * (98 + 91 + 87 + 83 + 81 + 76 + 74 + 68 + 59) = \frac{717}{9} \approx 79{,}67$$

Sind Werte wie hier einigermaßen fair verteilt, ist das arithmetische Mittel eine gute Möglichkeit, ein durchschnittliches Verhalten zu ermitteln. Manchmal gibt es allerdings sogenannte statistische Ausreißer, also einzelne Extremwerte, die einen Mittelwert extrem verzerren. Bei obigen Testergebnissen verändert sich der Mittelwert beispielsweise deutlich, wenn ein Test mit dem Ergebnis 0 Punkte hinzukommt:

$$\bar{x}_{arithm} = \frac{1}{10} * (98 + 91 + 87 + 83 + 81 + 76 + 74 + 68 + 59 + 0) = \frac{717}{10} \approx 71{,}17$$

Das Problem kann auch noch deutlicher ausfallen, etwa bei Durchschnittseinkommen. Nehmen Sie der Einfachheit halber an, ein kleines Unternehmen habe die Gehaltsstruktur eines DAX-Großkonzerns. Der CEO erhält ein Jahresgehalt von 1 Million Euro pro Jahr, fünf leitende Angestellte je 50.000, zehn Sachbearbeiter*innen je 40.000 und fünf Azubis je 15.000. Das ergibt das folgende durchschnittliche Jahresgehalt:

$$\bar{x}_{arithm} = \frac{1}{21} * (1.000.000 + 5 * 50.000 + 10 * 40.000 + 5 * 15.000) = \frac{1.725.000}{21} \approx 82142{,}86$$

Steigt das Gehalt des CEO auf 1,5 Millionen Euro, während die anderen Gehälter konstant bleiben, beträgt das Durchschnittsgehalt sogar etwa 105.952,38 €. Dies bildet die realen Verhältnisse nicht sinnvoll ab.

Um solche Statistikprobleme zu lösen, gibt es verschiedene Strategien:

- Ausreißer aus der Statistik nehmen – Ohne den CEO beträgt das Durchschnittsgehalt in dem fiktiven Minikonzern 36.250 €, was näher an einem für die meisten Angestellten typischen Wert liegt.
- Gruppierte Mittelwerte statt eines Gesamtmittelwerts berechnen – Mittelwerte sind innerhalb von Gruppen einigermaßen vergleichbarer Werte sinnvoller als über Werte mit massiven Proportionsunterschieden verteilt. In der Gehaltsstruktur eines Unternehmens ist es beispielsweise empfehlenswert, jeweils getrennte Durchschnitte für die Gehälter von Vorstand, leitenden Angestellten, Sachbearbeiter*innen und Azubis zu bilden. Diese Erkenntnis ist eine der Grundideen der Clustering-Algorithmen, die im Kapitel »Machine Learning« behandelt wird.
- Den Median statt des Mittelwerts betrachten

3.1.2.3 Median

Der Median ist der mittlere Wert, aber nicht der Mittelwert aus einer Gruppe von Werten: Wenn alle Werte geordnet nebeneinander geschrieben werden, ist der Median der Wert des mittleren Elements (oder bei einer geraden Anzahl von Werten der Mittelwert aus den beiden mittleren Elementen). Schematisch wird diese Definition wie folgt geschrieben:

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

Im Gehaltsbeispiel wäre der Median der Wert des 11. von 21 Werten insgesamt, also 40.000 (siehe Tabelle; Nr.11 (22/2)).

Nr.	Gehalt	Nr.	Gehalt
1	1.000.000,00 €	11	**40.000,00 €**
2	50.000,00 €	12	40.000,00 €
3	50.000,00 €	13	40.000,00 €
4	50.000,00 €	14	40.000,00 €
5	50.000,00 €	15	40.000,00 €
6	50.000,00 €	16	40.000,00 €
7	40.000,00 €	17	15.000,00 €
8	40.000,00 €	18	15.000,00 €
9	40.000,00 €	19	15.000,00 €
10	40.000,00 €	20	15.000,00 €
		21	15.000,00 €

Der Median ermöglicht mitunter sinnvolle statistische Aussagen,weil er ein für die Gesamtmenge typischer Wert ist.

3.1.2.4 Modus

Ein ähnlicher Wert ist der Modus $\bar{x}_d$,bei dem es sich um den am häufigsten vorkommenden Wert handelt. Dies wäre im vorliegenden Fall ebenfalls 40.000

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

3.1.2.5 Varianz

Wichtige Maße dafür, wie stark die betrachteten Werte voneinander abweichen, sind Varianz und Standardabweichung. Die Varianz wird folgendermaßen berechnet: Von jedem einzelnen Wert wird das arithmetische Mittel abgezogen, diese Differenzen werden quadriert (unter anderem, damit es nur positive Werte gibt) und zusammenaddiert, und schließlich wird dieser Wert durch n – 1 geteilt. Die Formel lautet also:

$$Var(X) = \frac{1}{n-1} * \sum_{i=1}^{n}(x_i - \bar{x})^2$$

3.1.2.6 Standardabweichung

Die Standardabweichung ist die Quadratwurzel aus der Varianz, die in vielen Fällen bevorzugt wird, denn erstens liegt der Wert durch die Aufhebung des Quadrierens näher an den tatsächlichen Abständen, und zweitens bleiben eventuelle Maßeinheiten auf diese Weise bestehen: Angenommen, die ursprünglichen Werte sind Abstände in Metern, dann wäre die Varianz in Quadratmetern, was überhaupt keinen Sinn ergäbe. Durch das Wurzelziehen ist die Einheit wieder Meter.

$$\sigma = \sqrt{Var(X)} = \sqrt{var(X) = \frac{1}{n-1} * \sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Hier sehen Sie die Berechnung der Varianz des Gehaltsverteilungsbeispiels, der Einfachheit halber in *Tausend €* mit $\bar{x} = 82{,}14286$

$$Var(X) = \frac{1}{21-1} * ((1000 - \bar{x})^2 + 5 * (50 - \bar{x})^2 + 10 * (40 - \bar{x})^2 + 5 * (15 - \bar{x})^2) \approx 44396{,}43$$

Die (aus einem etwas genaueren Näherungswert gezogene) Quadratwurzel dieser Varianz ist die Standardabweichung σ; sie beträgt etwa 210,7046 k€. Denken Sie daran, dass dieser Wert für die Einheit Tausend Euro berechnet wurde – die Standardabweichung für die realen Gehälter ist das Tausendfache dieses Wertes, also ungefähr 210.704,60 €.

Für Gruppen voneinander unabhängiger Zähl- oder Messwerte ist das arithmetische Mittel (mit den genannten Vorsichtsmaßnahmen gegen Ausreißer) oft die richtige Wahl. Es gibt jedoch andere Wertereihen, für die es nicht infrage kommt, sondern sogar falsche Ergebnisse liefert.

Haben Sie es etwa mit einer Reihe von n schwankenden prozentualen Veränderungen zu tun, liefert das arithmetische Mittel keinen Durchschnitt, der bei n Anwendungen auf den ersten Wert dasselbe Endergebnis liefert wie die Anwendung aller Einzelwerte.

Angenommen, ein Ausgangsbetrag von 1.000 € vermehrt sich im ersten Jahr um 5 %, schrumpft im zweiten Jahr um 1 % und steigt im dritten Jahr wieder um 6 %. Das Endergebnis erhalten Sie durch folgende Rechnung:

1000 * 1,05 * 0,99 *1,06 ≈ 1101,87

Das arithmetische Mittel aus den drei prozentualen Veränderungen

$$\bar{x}_{arithm} = \frac{1}{3} * (1{,}05 + 0{,}99 + 1{,}06) = \frac{3{,}1}{3} \approx 1{,}0\bar{3}$$

liefert wie vorhergesagt nicht das richtige Ergebnis:

$$1000 * 1{,}0\bar{3} * 1{,}0\bar{3} * 1{,}0\bar{3} \approx 1103{,}37$$

3.1.2.7 geometrischen Mittel

Das korrekte Ergebnis erhalten Sie mit dem geometrischen Mittel, das wie folgt definiert ist:

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

Im vorliegenden Fall wird das geometrische Mittel folgendermaßen berechnet:

$$\bar{x}_{geom} = \sqrt[3]{1{,}05 * 0{,}99 * 1{,}06} \approx 1{,}032864743$$

Mit einem etwas genaueren Näherungswert als hier gezeigt ergibt sich korrekterweise:

$$1000 * \bar{x}_{geom} * \bar{x}_{geom} * \bar{x}_{geom} \approx 1101{,}87$$

3.1.2.8 Harmonisches Mittel

Das harmonische Mittel ist ein Lageparameter der Statistik und kommt bei Verhältniszahlen zur Anwendung. Man berechnet mit ihm den Mittelwert der Menge dieser Zahlen. Als Verhältniszahlen werden Brüche bezeichnet, die eine Beziehung widerspiegeln. Also zum Beispiel Studenten pro Einwohner oder Preis pro Quadratmeter.

3.1.2.8.1 Harmonisches Mittel Beispiel

Angenommen du machst einen Wochenendausflug und fährst eine dreiteilige Strecke. Der erste Streckenabschnitt ist 50 km lang und du fährst 150 km/h. Der zweite Streckenabschnitt ist 60 km lang und der dritte 90 km. Auf den letzten beiden Strecken fährst du 120km/h bzw. 90 km/h.

Streckenabschnitt	Strecke in km	Geschwindigkeit in km/h	Zeit in h
1	50	150	⅓
2	60	120	½
3	90	90	1

3.1.2.8.2 Harmonisches Mittel berechnen

Jetzt sollst du die Geschwindigkeit berechnen, mit der du durchschnittlich unterwegs warst. Da es sich bei der Geschwindigkeit ja um eine Verhältniszahl handelt, nämlich Kilometer pro Stunde, geht das mit dem harmonischen Mittel.

3.1.2.8.3 Harmonisches Mittel Formel

Die allgemeine Formel zur Berechnung sieht so aus:

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

$$\frac{\text{Die Anzahl der Ausprägungen}}{\text{Die Summe der Kehrwerte}}$$

3.1.2.8.4 Harmonisches Mittel Anwendung

Um die Durchschnittsgeschwindigkeit zu berechnen müssen wir unsere Formel von oben leicht abwandeln:

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

Du teilst also die Summe der Länge der Teilstrecke durch die Summe der Quotienten aus der Teilstrecke und der Geschwindigkeit der Teilstrecke. In unserem Beispiel rechnen wir also 50 plus 60 plus 90 geteilt durch 50 durch 150 plus 60 durch 120 plus 90 durch 90. Als Ergebnis erhalten wir eine Durchschnittsgeschwindigkeit von 109,09 km/h.

$$\bar{x}_{harm} = \frac{\sum_{i=1}^{n} g_i}{\sum_{i=1}^{n} \frac{g_i}{x_i}} = \frac{50 + 60 + 90}{\frac{50}{150} + \frac{60}{120} + \frac{90}{90}} = 109{,}09$$

3.1.2.9 Harmonisches Mittel arithmetisches Mittel Unterschied

Wenn du die Durchschnittsgeschwindigkeit mit dem normalen arithmetischen Mittel berechnest, würdest du eine falsche Lösung erhalten, weil du nicht berücksichtigst, dass du die verschiedenen Geschwindigkeiten ja unterschiedlich lange fährst. Es gibt aber einen Trick wie du auch mit dem arithmetischen Mittel auf die richtige Lösung kommst und zwar indem du die Geschwindigkeiten mit den Zeiten gewichtest. Die Formel dazu sieht so aus:

3.1.2.9.1 Gewichtetes arithmetisches Mittel

$$\bar{x}_{gew.arithm} = \frac{\sum_{i=1}^{n} h_i * x_i}{\sum_{i=1}^{n} h_i}$$

h_i steht dabei für die Zeit, die du brauchst um den jeweiligen Streckenabschnitt zurückzulegen und x_i steht wieder für die Geschwindigkeit. In unserem Beispiel würde die Rechnung also so aussehen:

$$\bar{x}_{gew.arithm} = \frac{\sum_{i=1}^{n} h_i * x_i}{\sum_{i=1}^{n} h_i} = \frac{\frac{1}{3} * 150 + \frac{1}{2} * 120 + 1 * 90}{\frac{1}{3} + \frac{1}{2} + 1} = 109{,}09\ km/h$$

Wie du siehst, kannst du also auch mit dem gewichteten arithmetischen Mittel auf das richtige Ergebnis kommen!

Jetzt weißt du wie du das **harmonische Mittel** berechnest und dass du mit dem **gewichteten arithmetischen Mittel** auf **dasselbe Ergebnis** kommst!

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

4 Kombinatorik

Ziehen aus einer Urne? In der Kombinatorik erklären wir dir die wichtigsten Formeln dazu.

4.1 Urnenmodell

Ein Urnenmodell dient in der Stochastik zur vereinheitlichten Darstellung und Modellierung von Zufallsexperimenten.

Du kannst mithilfe eines Urnenmodells aber nicht nur die Frage beantworten „Wie hoch ist die Wahrscheinlichkeit zwei weiße Kugeln zu ziehen?“, sondern zum Beispiel auch die Anzahl an Möglichkeiten bestimmen in welcher Reihenfolge die Kugeln gezogen werden. So lassen sich beispielsweise Alltagssituationen abbilden und man kann so die Wahrscheinlichkeit berechnen für die verschiedenen Szenarien.

Das Urnenmodell verwendet dazu einen Behälter, zum Beispiel eine Kiste, in der sich verschiedene, sagen wir schwarze und weiße Kugeln befinden. Nun werden aus der Kiste, ohne hineinzusehen, Kugeln gezogen und es wird notiert, ob diese schwarz oder weiß sind. Dabei gibt es verschiedene Varianten wie dieses Zufallsexperiment durchgeführt wird. Man unterscheidet, ob eine gezogene Kugel wieder zurückgelegt wird und ob die Reihenfolge eine Rolle spielt oder nicht.

4.2 Kombinatorik Urnenmodell

Generell unterschiedet man in der Kombinatorik zwischen Stichproben mit Reihenfolge, die dann Variation genannt werden, und Stichproben ohne Reihenfolge, die Kombination genannt werden. Je nachdem, ob man die Kugeln dann noch zurück legt oder nicht, ergeben sich dann die verschiedenen Urnenmodelle.

4.2.1 Ziehen ohne Zurücklegen ohne Reihenfolge

So los geht es mit Kombination ohne Wiederholung. Du hast es also mit dem Szenario zu tun, dass die Reihenfolge der Ergebnisse des Zufallsexperimentes keine Rolle spielt und das Ergebnis nicht erneut eintreten kann, wenn es bereits aufgetreten ist.

Schauen wir uns das Ganze gleich anhand eines praktischen Beispiels an. Stell dir vor du hast eine Kiste mit 8 schwarzen und 4 weißen Kugeln. Nun nimmst du nacheinander 4 Kugeln aus der Kiste, ohne sie danach zurückzulegen.

Jetzt möchtest du wissen, wie viele mögliche Ergebnisse du bei dieser Ziehung erhalten kannst. Das bestimmst du mit Hilfe des ***Binomialkoeffizienten***. Hier zur Wiederholung nochmal die Formel:

$$\binom{n}{k} = \frac{n!}{k! * (n-k)!}$$

n steht hierbei für die Anzahl an Elementen insgesamt und klein k für die Anzahl an Ziehungen. Wir rechnen also:

$$\binom{12}{4} = \frac{12!}{4! * (12-4)!} = 495$$

Es gibt also 495 Möglichkeiten die Kugeln aus der Urne zu ziehen.

4.2.2 Wahrscheinlichkeit ohne Zurücklegen ohne Reihenfolge

Als nächstes möchtest du die Wahrscheinlichkeit bestimmen, genau eine schwarze Kugel zu ziehen. Um das zu berechnen, musst du wissen, dass diesem Zufallsexperiment die hypergeometrische

Verteilung zugrunde liegt. Mithilfe der Formeln der Verteilung kannst du diese Aufgabe lösen. Genauer gesagt verwenden wir die Funktion für die Dichte der ***hypergeometrischen Verteilung***, denn diese Wahrscheinlichkeitsfunktion gibt ja die Wahrscheinlichkeit im diskreten Fall dafür an, genau einen Wert x zu erhalten. Die Formel der Funktion:

$$f(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

N ist dabei die Anzahl der Elemente insgesamt, bei uns gilt also N ist gleich 12. M gibt die Anzahl derjenigen Elemente an, die als „Erfolg" gesehen werden. Da wir uns ja für die schwarzen Kugeln interessieren, gilt M gleich 8. Klein n steht für die Anzahl an Elementen, die für das Zufallsexperiment gezogen werden, bei uns ist also klein n gleich 4.

Wenn du nun wissen möchtest mit welcher Wahrscheinlichkeit genau eine schwarze Kugel gezogen wird, musst du einfach die Wahrscheinlichkeit für x gleich 1 berechnen. Wenn wir alles einsetzen, erhalten wir folgende Berechnung:

$$f(1) = \frac{\binom{8}{1}\binom{12-8}{4-1}}{\binom{12}{4}} = \frac{8*4}{495} = 0{,}0646$$

Die Wahrscheinlichkeit genau eine schwarze Kugel zu ziehen liegt also bei **ungefähr 6,46%**.

Hier findest du nochmal die wichtigsten Formeln für Ziehen ohne Zurücklegen ohne Reihenfolge im Überblick:

- Binomialkoeffizient (Anzahl an Möglichkeiten berechnen)
- Wahrscheinlichkeitsfunktion (Wahrscheinlichkeit genau x schwarze Kugeln zu ziehen)
- Verteilungsfunktion (Wahrscheinlichkeit weniger oder genau x schwarze Kugeln zu ziehen)

$$F(x) = P(X \leq x) = \sum_{x_i=1}^{x} f(x_i)$$

Die Wahrscheinlichkeit dafür, dass die Zufallsgröße X kleiner gleich x ist, entspricht der Summe aller Einzelwahrscheinlichkeiten x_i, solange x_i kleiner gleich x ist.

4.2.3 Ziehen ohne Zurücklegen mit Reihenfolge

Jetzt weißt du wie du Aufgaben zum Ziehen aus der Urne ohne Zurücklegen und ohne Beachtung der Reihenfolge lösen kannst. Bisher hat es keinen Unterschied gemacht, in welcher Reihenfolge die Kugeln gezogen wurden, also zum Beispiel erst zwei schwarze und dann zwei weiße oder anders herum. Nun betrachten wir eine Variation ohne Wiederholung, also den Fall, dass die Reihenfolge eine Rolle spielt.

4.2.3.1 Urnenmodell ohne Zurücklegen mit Reihenfolge

In diesem Fall legen wir die Kugeln also nicht zurück und die Reihenfolge ist entscheidend für das Ergebnis. Ein anschauliches Beispiel hierfür ist, wie viele Möglichkeiten es gibt die ersten drei Plätze bei einem Beerpong-Turnier mit 15 teilnehmenden Gruppen zu besetzen. Hier macht es nämlich natürlich einen Unterschied, ob eine Gruppe auf dem ersten oder auf dem dritten Platz landet.

Die Formel, um die Anzahl an Möglichkeiten zu berechnen, können wir uns ganz einfach selbst logisch herleiten.

Wir haben 15 Teams, die den ersten Platz belegen können. Nachdem dieser vergeben wurde, bleiben noch 14 Teams, die eine Chance auf den zweiten Platz haben. Danach bleiben schließlich noch 13 Teams, die den dritten Platz belegen können. Um die Gesamtanzahl an Möglichkeiten zu berechnen, rechnest du also 15 mal 14 mal 13 gleich 2.730 Möglichkeiten.

4.2.3.2 Formel Anzahl an Möglichkeiten

Die allgemeine Formel lautet bei Ziehungen ohne Zurücklegen mit Beachtung der Reihenfolge N Fakultät geteilt durch N minus k Fakultät.

$$\frac{N!}{(N-k)!}$$

Groß N steht dabei für die Anzahl an Elementen insgesamt, in unserem Fall sind das die 15 Teams, und klein k steht für Anzahl an Ziehungen, in unserem Fall gilt also k gleich 3 da wir ja die ersten 3 Plätze belegen möchten. Wenn wir diese Angaben einsetzen, erhalten wir auch wieder genau die 2.730 Möglichkeiten.

Das war auch schon alles, was du zu Variationen ohne Zurücklegen wissen musst!

4.2.4 Ziehen mit Zurücklegen ohne Reihenfolge

Genau wie bei den Ziehungen ohne Zurücklegen bietet sich das Urnenmodell an, um das Vorgehen verständlich zu erklären. Gehen wir davon aus, dass wir eine Kiste mit 8 schwarzen und 4 weißen Kugeln haben. Wir ziehen daraus wieder, ohne hineinzusehen, 4 Kugeln, nur dass wir sie diesmal nach jedem Zug wieder hineinlegen.

Es befinden sich also nach jedem Zug gleich viele Kugeln in der Urne. Jetzt möchtest du wissen, wie viele mögliche Ergebnisse du bei den 4 Ziehungen erzielen kannst, zum Beispiel nur weiße Kugeln, nur schwarze Kugeln, 2 weiße und 2 schwarze und so weiter. Du hast es also mit einem Urnenmodell mit Zurücklegen ohne Reihenfolge zu tun. Wie du jetzt bereits weißt, spricht wann von Kombinationen, wenn die Reihenfolge keine Rolle spielt.

4.2.4.1 Wahrscheinlichkeit Ziehen mit Zurücklegen ohne Reihenfolge

Du kannst die Aufgaben zu diesem Szenario des Zufallsexperiments nun mithilfe des Binomialkoeffizienten und der Binomialverteilung lösen.

Um die Anzahl an Möglichkeiten zu berechnen benötigst du eine leicht abgewandelte Form des Binomialkoeffizienten:

$$\binom{N+k-1}{k} = \frac{(N+k-1)!}{(N-1)! * k!}$$

N steht dabei für die Anzahl an Kugeln insgesamt und klein k für die Anzahl an Ziehungen. Wenn wir die gegebenen Werte einsetzen, erhalten wir also:

$$\binom{12+4-1}{4} = \frac{(12+4-1)!}{(12-1)! * 4!} = 1365$$

Es gibt also 1365 verschiedene mögliche Ergebnisse.

Als nächstes möchtest du noch die Wahrscheinlichkeit bestimmen, genau eine schwarze Kugel zu ziehen. Dazu musst du wissen, welche Verteilung diesem Zufallsexperiment zugrunde liegt. Bei Ziehungen mit Zurücklegen und ohne Reihenfolge ist das die Binomialverteilung.

Um die Aufgabe zu lösen, benötigst du also die *Wahrscheinlichkeitsfunktion* der Binomialverteilung. Zur Wiederholung hier noch einmal die Formel:

$$f(k) = P(X = k) = \binom{n}{k} * p^k * (1 - p)^{n-k}$$

Klein n steht dabei für die Anzahl der Ziehungen. Für die Anzahl an Treffern steht k. Klein p steht für die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen. Da 8 von 12 Kugeln schwarz sind, gilt p=$\frac{2}{3}$. Da wir nach jedem Zug die Kugel wieder zurück legen bleibt diese Wahrscheinlichkeit immer gleich. Jetzt können wir alle Werte einsetzen:

$$f(1) = \binom{n}{k} * \left(\frac{2}{3}\right)^1 * (1 - \frac{2}{3})^{4-1} =$$

$$f(k) = P(X = k) = \binom{4}{1} * \left(\frac{2}{3}\right)^1 * \left(1 - \frac{2}{3}\right)^{4-1} = 4 * \frac{2}{3} * \left(\frac{1}{3}\right)^3 = 0{,}09877$$

Die Wahrscheinlichkeit genau eine schwarze Kugel zu ziehen liegt also bei ungefähr 9,9%.

Zusammenfassend solltest du dir merken, dass Zufallsexperimente mit Ziehungen mit Zurücklegen und ohne Reihenfolge einer Binomialverteilung folgen. Das heißt, du musst die Formeln der Binomialverteilung zur Lösung solcher Aufgaben verwenden.

4.2.5 Ziehen mit Zurücklegen mit Reihenfolge

Aber wie sieht es aus bei Ziehungen mit Zurücklegen mit Reihenfolge? Auch das ist kein Hexenwerk, wenn du weißt welche Formel du bei Ziehungen mit Zurücklegen unter Beachtung der Reihenfolge verwenden musst. Zuerst ist es wichtig, dass du dir erst noch einmal klarmachst, um welches Urnenmodell es sich handelt.

4.2.5.1 Variation mit Wiederholung

Wir betrachten also Variationen, genauer gesagt Ziehungen mit Zurücklegen, bei denen die Reihenfolge einen Unterschied macht. Ein anschauliches Beispiel hierfür ist der Code eines Fahrradschlosses. Die Reihenfolge der Zahlen machen einen Unterschied, allerdings kann jede Zahl beliebig oft vorkommen.

4.2.5.2 Wahrscheinlichkeit Ziehen mit Zurücklegen mit Reihenfolge

Gehen wir davon aus, du hast die 5-stellige Kombination deines Fahrradschlosses vergessen. Jede Zahl könnte eine Ziffer zwischen 1 und 6 sein. Wie viele Möglichkeiten kannst du ausprobieren?

Für jede der 5 Stellen der Kombination gibt es 6 Möglichkeiten. Insgesamt gibt es also 6 hoch 5 gleich 7.776 mögliche Kombinationen für das Zahlenschloss.

$$N^k$$

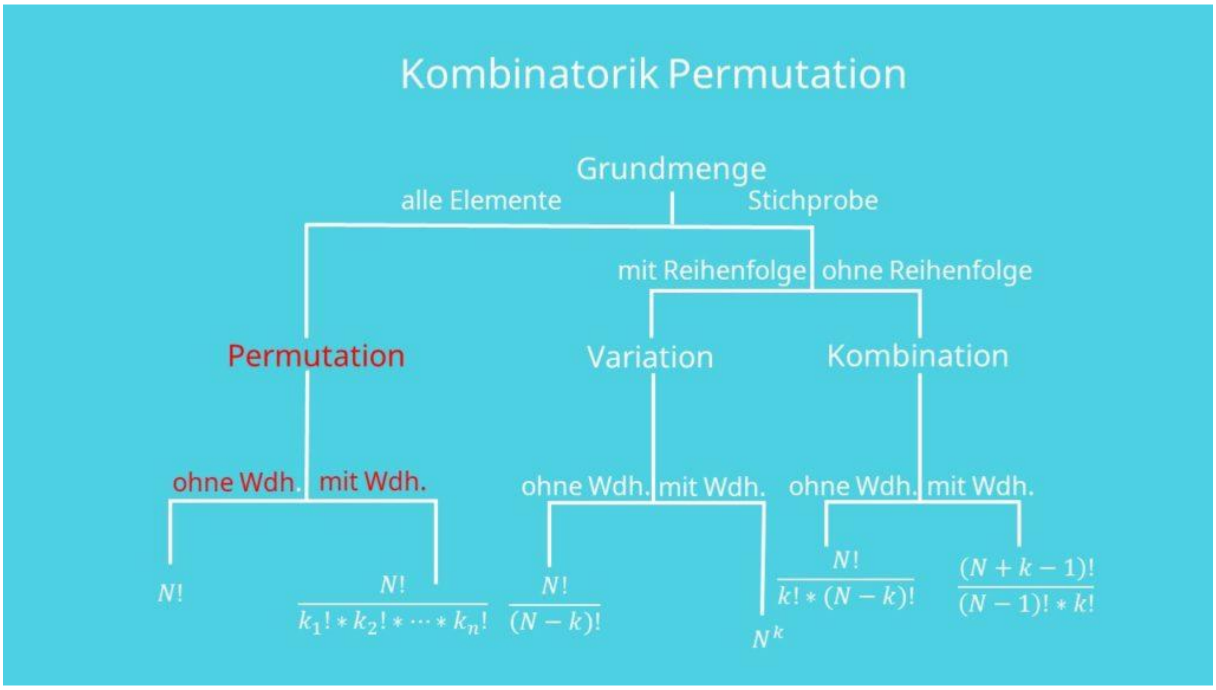
Groß N steht dabei wieder für die Anzahl an Elementen, aus denen gezogen wird, in unserem Fall also die 6 möglichen Ziffern, und klein k steht für die Anzahl der Ziehungen, die in diesem Fall den 5 Stellen der Kombination entsprechen.

4.2.6 Permutation Definition

Als Permutation wird in der Kombinatorik eine mögliche Anordnung von Objekten bezeichnet. Je nachdem ob alle Objekte unterscheidbar voneinander sind oder nicht, handelt es sich um eine Permutationen mit Wiederholung oder ohne Wiederholung.

4.2.7 Kombinatorik Permutation

Wie auch bei den Variationen und den Kombinationen, unterscheidet man also auch bei den Permutationen zwischen solchen ohne und solchen mit Wiederholung. Im Folgenden findest du eine Einordnung von Permutationen in eine Übersicht aller Formeln der Kombinatorik.



4.2.7.1 Unterschied Permutation Kombination

Generell unterscheidet man in erster Linie, ob man alle Objekte oder nur einen Teil davon betrachtet. Gehen wir davon aus, dass nur eine Teilmenge der Grundgesamtheit für die Berechnung der Möglichkeiten relevant ist, so spricht man von Kombinationen beziehungsweise Variationen. Bei einer Kombination ist im Gegensatz zur Variation ist die Reihenfolge der Anordnung nicht relevant. Trifft man dagegen keine Auswahl, so berechnet man die Möglichkeiten die Elemente anzuordnen mithilfe von Permutationen. Permutationen ähneln grundsätzlich sehr stark den Variationen. Der einzige Unterschied ist, dass bei Permutationen die Besonderheit N=k gilt. Das heißt dass aus insgesamt N Elementen alle Elemente gezogen werden und nicht nur die Teilmenge relevant ist.

4.2.8 Permutation mit Wiederholung

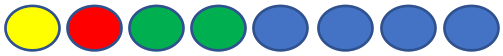
Betrachten wir zuerst Permutationen mit Wiederholung. Jetzt fragst du dich vielleicht, wie es eine Wiederholung geben kann, wenn alle Elemente auf einmal gezogen werden.

Man spricht von Permutationen mit Wiederholung, wenn es Elemente in der Ausgangsmenge gibt, die **nicht voneinander unterscheidbar** sind, also zum Beispiel Kugeln derselben Farbe. Anhand eines Beispiels wird das ganze gleich verständlicher.

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

4.2.8.1 Permutation Beispiel

Stell dir vor, du hast 8 Kugeln. Eine davon ist gelb, eine ist rot, 2 sind grün und 4 sind blau. Nun sollst du herausfinden, wie viele Möglichkeiten es gibt diese Kugeln anzuordnen. Man kann also jeweils die beiden grünen und die 4 blauen Kugeln nicht voneinander unterscheiden.



4.2.8.2 Permutation Formel

Deshalb muss man die musst du die Formel der N Fakultät, leicht abwandeln, indem du sie durch das Produkt der Fakultäten der Häufigkeiten jedes Elements teilst. Allgemein sieht die Formel bei Permutationen mit Wiederholung dann so aus:

$$\frac{N!}{k_1! * k_2! * \ldots * k_n!}$$

4.2.8.3 Permutation berechnen

Setzten wir die Zahlen unseres Beispiels ein, so erhalten wir:

$$\frac{8!}{1! * 1! * 2! * 4!} = 840$$

Es gibt also 840 Möglichkeiten, die Kugeln anzuordnen.

4.2.9 Permutation ohne Wiederholung

Während es bei Permutationen mit Wiederholung Elemente in der Ausgangsmenge gibt, die nicht voneinander unterscheidbar sind, unterscheiden sich im Fall ohne Wiederholung alle Elemente voneinander. Das heißt, dass jedes Objekt tatsächlich einzigartig ist bezüglich seiner Merkmalsausprägungen.

4.2.9.1 Permutation Beispiel

Ein Beispiel hierfür wäre, dass 10 Studenten den Vorlesungssaal verlassen. Nun sollst du berechnen, wie viele Reihenfolgen dabei möglich sind.

4.2.9.2 Permutation Formel

Allgemein lautet die Formel zur Berechnung der Anzahl der Möglichkeiten bei Permutationen ohne Wiederholung ganz einfach N Fakultät:

$$N!$$

Einfach gesagt multipliziert man also einfach die Anzahl der verbleibenden Möglichkeiten auf.

4.2.9.3 Permutation berechnen

Für den ersten Student, der die Vorlesung verlässt, gibt es noch 10 Möglichkeiten. Für den zweiten schon nur noch 9 und so weiter. Insgesamt gibt also 10 mal 9 mal 8 mal 7 etc., also 10 Fakultät Möglichkeiten. Das sind insgesamt 3.628.800 mögliche Reihenfolgen der Studenten!

10 * 9 * 8 * 7 * 6 * 5 * 4 * 3 * 2 *1 = 10! = 3.628.800

So, das wars auch schon zu Permutationen! Zusammenfassend musst du dir also nur merken, dass Permutationen eine Art **Sonderform** der **Variationen** mit **N=k** darstellen. Im Falle einer **Wiederholung** ist die allgemeine Formel zur Berechnung der Möglichkeiten:

$$\frac{N!}{k_1! * k_2! * \ldots * k_n!}$$

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de

Bei *Permutationen ohne Wiederholung* kannst du die Anzahl an Möglichkeiten ganz einfach mit N Fakultät (N!) berechnen.

5 Wahrscheinlichkeitsverteilungen

Wahrscheinlichkeitsverteilungen sind spezielle Ansätze bestimmte Probleme der Wahrscheinlichkeitsrechnung zu lösen und basieren auf den Grundlagen der Stochastik.
Das sind die wichtigsten Wahrscheinlichkeitsverteilungen :

- Bernoulliverteilung https://de.wikipedia.org/wiki/Bernoulli-Verteilung
- Binomialverteilung https://de.wikipedia.org/wiki/Binomialverteilung
- Hypergeometrische Verteilung https://de.wikipedia.org/wiki/Hypergeometrische_Verteilung
- Geometrische Verteilung https://de.wikipedia.org/wiki/Geometrische_Verteilung
- Poisson Verteilung https://de.wikipedia.org/wiki/Poisson-Verteilung
- Diskrete Verteilung https://de.wikipedia.org/wiki/Diskrete_Wahrscheinlichkeitsverteilung
- Stetige Verteilung https://de.wikipedia.org/wiki/Stetige_Wahrscheinlichkeitsverteilung
- Exponentialverteilung https://de.wikipedia.org/wiki/Exponentialverteilung
- Chi Quadrat Verteilung https://de.wikipedia.org/wiki/Chi-Quadrat-Verteilung
- t Verteilung https://de.wikipedia.org/wiki/Studentsche_t-Verteilung

Quellen: https://studyflix.de/statistik/stochastik-1931
https://de.serlo.org/mathe/1875/vierfeldertafel
Daten- und Prozessanalyse, ISBN 978-3-8362-8112-6
www.lektorat-plus.de