
ORIENTATE Workshopmaterial

Glossar statistischer Grundbegriffe

D Datenerhebung

Datenerhebung ist die systematische und zielgerichtete Tätigkeit zur Beschaffung von Informationen, die die Grundlage für ein empirisches Forschungsprojekt bilden. Diese Informationen dienen dazu, entscheidungsrelevante Daten zu generieren und die Forschungsziele zu erreichen. Folgende Methoden sind Beispiel einer Datenerhebung: Fragebogen, Interviews, Fokusgruppen uvm.

E Einheit

Die statistische Einheit bezieht sich auf die Elemente oder Beobachtungseinheiten, für die Individualdaten erhoben werden, während sich der Begriff "Erhebungseinheit" auf die Organisation oder Institution bezieht, die Informationen in einer statistischen Erhebung bereitstellt. Beispielsweise sind statistische Einheiten Personen, Kommunen, Religionsgruppen usw.

G Generalisierung

In der Forschung bezeichnet "Generalisierung" den Prozess der Übertragung von Ergebnissen einer Stichprobe auf die gesamte Grundgesamtheit (siehe Population). Das bedeutet, dass die Schlussfolgerungen, die aus den Daten einer begrenzten Gruppe von Studienteilnehmern gezogen wurden, auf eine breitere Population angewendet werden, um zu allgemeinen Schlussfolgerungen zu gelangen. Dieses Verfahren ist sinnvoll, wenn die Stichprobe repräsentativ (siehe repräsentativ) für die Grundgesamtheit ist.

H Hypothese

Statistische Hypothesen sind vorab formulierte Erwartungen oder Annahmen darüber, wie sich Variablen (Merkmalen) in einem Datensatz voneinander unterscheiden oder miteinander in Beziehung stehen könnten. Sie dienen dazu, Vermutungen oder Vorhersagen über das Auftreten bestimmter Merkmale oder Phänomene in den Daten zu machen. Die Formulierung statistischer Hypothesen ist ein wichtiger Schritt, der immer vor Beginn einer Untersuchung erfolgt. Ein Beispiel wäre "Je mehr Fenster ein Büro hat, desto produktiver sind die Mitarbeiter."

I Inferenzstatistik

„Inferenzstatistik“ wird oft auch als „schließende Statistik“ bezeichnet. Der Begriff bezeichnet den Rückschluss von einer Stichprobe (siehe Stichprobe) auf Werte in der Population (siehe Population) mittels Generalisierungen (siehe Generalisierung). Beispielsweise kann die Frage beantwortet werden, ob Basketballspieler größer sind als der Durchschnitt der männlichen Bevölkerung.

Interquartilbereich

Der Interquartilbereich (IQR) ist der Abstand zwischen dem unteren Quartil (siehe Quartil) (Q1) und dem oberen Quartil (Q3) einer Datenverteilung. Er gibt an, wie weit die mittleren 50% der Daten im Datensatz voneinander entfernt sind, und hilft bei der Beschreibung der Streuung der Daten. IQR ist ein ausreißerresistentes Streuungsmaß, das hilft, die Variabilität innerhalb der mittleren 50% der Daten zu verstehen. Beispiel: in einem Ort gibt es 10 Kinder im Alter von 1, 3, 5, 7, 8, 9, 11, 12, 14 und 16 Jahren: Interquartilsabstand = 3. Quartil - 1. Quartil = 12 Jahre - 5 Jahre = 7 Jahre.

K Kausalität

Ein kausaler Zusammenhang liegt vor, wenn ein Effekt (eine Wirkung) auf die vermutete Ursache zurückgeführt werden kann. Zum Beispiel, wenn höhere Temperaturen im Sommer zu einer erhöhten Eiscreme-Konsum führen.

Korrelation

Die Korrelation gibt an, inwieweit zwei Variablen so miteinander zusammenhängen, dass sie sich in einem festen Verhältnis zueinander verändern. Korrelation wird verwendet, um einfache Beziehungen zwischen Variablen zu beschreiben, ohne zu bestimmen, welche die Ursache und welche die Wirkung ist. Zum Beispiel, wenn eine positive Korrelation zwischen dem Verzehr von Eiscreme und der Anzahl der Badeunfälle im Sommer gefunden werden. Hier wird erst nur eine Korrelation beobachtet, da es ein Fehler wäre zu schlussfolgern, dass der Eiscreme-Verzehr Badeunfälle verursacht. In Wirklichkeit könnte die höhere Temperatur im Sommer sowohl den Eiscreme-Konsum als auch das Schwimmen im Freien erhöhen, was zu der beobachteten Korrelation führt.

Korrelationskoeffizient

Der Korrelationskoeffizient in der Statistik ist eine Zahl zwischen -1 und +1, die angibt, wie stark und in welche Richtung zwei Variablen miteinander verbunden sind. Ein Wert von -1 bedeutet, dass zwischen den beiden Variablen ein negativer Zusammenhang besteht. Bei einem Wert von +1 geht man von einem positiven Zusammenhang aus. Bei einem Wert von 0 besteht kein Zusammenhang zwischen den Variablen.

M Median

Der Median einer Gruppe von Werten ist definiert als der Wert, der größer ist als genau die Hälfte der Werte in der Gruppe. Der Median ist eine einfache Methode, um die Mitte der Werte in der Gruppe zu kennzeichnen, da genau die Hälfte der Werte kleiner und die andere Hälfte der Werte größer als der Median ist. Z.B., wenn wir die Reihe „1“, „5“ und „6“ haben ist der Median „5“.

Merkmalsarten

Die Arten von Merkmalen, die in der Statistik erfasst werden, lassen sich grob in zwei Kategorien einteilen: qualitative Merkmale und quantitative Merkmale. Qualitative Merkmale sind Merkmale, die eine Beschreibung (Qualität) darstellen und häufig mit Worten beschrieben werden (z.B., wie die Erfahrung von Personen ist, die Eiscreme essen). Quantitative Merkmale hingegen stellen ein Ausmaß dar und werden meist in Zahlen gemessen (z.B., wie viel Eiscreme eine Person in einem Sommer gegessen hat). Dies ermöglicht eine numerische Erfassung und Analyse dieser Merkmale.

Merkmalsausprägung

Ein statistisches Merkmal ist eine charakteristische Eigenschaft eines Objekts, die Gegenstand einer statistischen Untersuchung ist. Die verschiedenen Ausprägungen oder Varianten dieses Merkmals werden als Merkmalsausprägungen bezeichnet. Zum Beispiel hat das Merkmal „Alter“ die Merkmalsausprägungen „20“, „18“ und „37“.

Merkmalsträger

Ein Merkmalsträger, auch statistische Einheit genannt, ist das Grundelement der Statistik. Ein Merkmalsträger enthält Informationen oder Eigenschaften, die für eine statistische Untersuchung von Bedeutung sind. Es kann sich dabei z. B. um eine Person, ein Produkt, ein Unternehmen oder ein anderes Objekt handeln, das im Rahmen der statistischen Analyse betrachtet wird.

Messinstrument

Ein Messinstrument ist in der sozial-empirischen Forschung ein Mittel, mit dem Beobachtungen in Daten (Messwerte) überführt werden. Beispiele sind Fragebogen sowie Leitfaden für Fokusgruppen und Interviews. Bei technischen Anwendungen sowie in den Naturwissenschaften werden zur Datenerhebung insbesondere Messgeräte genutzt.

Metrisch

Metrisch skalierte Merkmale können der Größe nach geordnet werden, und es ist möglich, sinnvolle Summen oder Differenzen zwischen den Werten zu bilden, da die Abstände zwischen den Merkmalswerten gleich groß sind. Ein Beispiel für ein metrisch skaliertes Merkmal wäre das Alter in Jahren, da es klare Abstände zwischen den Werten gibt und Additionen und Differenzen eine sinnvolle Bedeutung haben.

Mittelwert (arithmetisches Mittel)

Die Summe aller Werte in einer Datenmenge geteilt durch die Anzahl der Werte (Durchschnitt). Z.B., wenn wir die Reihe „1“, „5“ und „6“ haben ist der Mittelwert „4“.

Modalwert

Der Wert, der in einer Verteilung am häufigsten vorkommt, wird Modalwert genannt. Um den Modalwert zu bestimmen, wird die Anzahl der einzelnen Werte in der Stichprobe gezählt. Z.B., wenn wir die Reihe „1“, „1“, „5“ und „6“ haben ist der Modalwert „1“.

N Nominal

Nominal skalierte Merkmale sind Merkmale, die keine natürliche Rangfolge haben und nicht in eine Reihenfolge gebracht werden können. Zum Beispiel sind „männlich“ und „weiblich“ nominale Merkmale, bei denen es keine Hierarchie gibt. Wenn es nur zwei Ausprägungen gibt, werden die Merkmale als binär oder dichotom bezeichnet.

Normalverteilung

Die Normalverteilung ist eine Art der Datenverteilung, die einer Glockenkurve ähnelt und nach dem deutschen Mathematiker Carl Friedrich Gauß (Gauß-Verteilung) benannt wurde. Sie ist ein in der Statistik häufig verwendetes Modell. Bei der Normalverteilung sind die Daten symmetrisch um den Mittelwert (siehe Mittelwert) verteilt, das heißt, Mittelwert und Median (siehe Median) sind gleich. Dieses Verteilungsmodell kann oft auf Daten aus einer großen Grundgesamtheit (siehe Grundgesamtheit) angewendet werden (z.B. Durchschnittsgröße der Menschen in Deutschland).

Nullhypothese

Die Nullhypothese ist die Behauptung, dass es in der Grundgesamtheit keinen Effekt gibt. Wenn die Stichprobe genügend Beweise gegen diese Behauptung liefert ($p\text{-Wert} \leq \alpha$, wobei α das Signifikanzniveau ist), so kann die Nullhypothese verworfen werden. Andernfalls wird die Nullhypothese nicht abgelehnt. Mit anderen Worten, wenn die Wahrscheinlichkeit hoch ist, dass die beobachteten Unterschiede in den Daten rein zufällig sind ($p\text{-Wert} > \alpha$), bleibt die Nullhypothese bestehen. Z.B. Es gibt keinen Zusammenhang zwischen höheren Temperaturen im Sommer und dem erhöhten Eiscreme-Konsum.

O Ordinal

Ordinalskalierte Merkmale können der Größe nach geordnet werden, aber es ist nicht immer sinnvoll, Summen oder Differenzen zwischen den Werten zu bilden. Beispielsweise sind Schulnoten ein ordinalskaliertes Merkmal, da „Eins“ besser als „Zwei“ ist, aber die Differenz zwischen den Noten hat keine klare quantitative Bedeutung, da die Abstände zwischen den Werten variabel sind.

Operationalisierung

Operationalisierung in der Statistik bezieht sich auf die Formulierung von Regeln, die festlegen, in welchem Umfang der durch einen Begriff beschriebene Sachverhalt vorliegt oder wie er beobachtbar ist. Dies beinhaltet die Verknüpfung abstrakter Begriffe mit konkreten, messbaren Sachverhalten und stellt sicher, dass Forschende in der Lage sind, Daten zu erheben und Sachverhalte zu quantifizieren. Um das Verhalten „Eiscreme essen“ zu operationalisieren könnte man z.B. festlegen, dass jede Kugel eines Eis oder nur jedes verkaufte Eis gezählt wird.

P p-Wert

Der p-Wert ist ein Wahrscheinlichkeitswert in der Statistik, der angibt, wie stark die Daten gegen eine aufgestellte Hypothese (Nullhypothese) sprechen. Wenn der p-Wert sehr klein ist, bedeutet dies, dass ein starker Beweis gegen die Nullhypothese (siehe Nullhypothese) vorliegt. Wenn der p-Wert unter einem bestimmten Grenzwert liegt, bezeichnet man die Ergebnisse als „statistisch signifikant“ (siehe Signifikanz). Z.B.: Der p-Wert von 0,03 liegt unter dem üblichen Signifikanzniveau von 0,05. Daher gibt es statistische Evidenz dafür, die Nullhypothese abzulehnen. Das bedeutet, dass es wahrscheinlich ist, dass ein Zusammenhang zwischen höheren Temperaturen im Sommer und dem erhöhten Eiscreme-Konsum besteht.

Population

Der Begriff Population oder Grundgesamtheit bezieht sich auf die gesamte Gruppe von Personen oder Objekten, über die wir mehr erfahren möchten. Zum Beispiel könnte die Population alle Menschen in einer bestimmten Altersgruppe oder Region umfassen (z. B. alle Kinder im Alter von 4 bis 6 Jahren in Deutschland), wenn Fragen zur Gesundheit und zum Wohlbefinden dieser Gruppe gestellt werden. Die Grundgesamtheit wird in der Regel mit dem Buchstaben N angegeben.

Q Quartil(e)

Quartile, auch als Angelpunkte bezeichnet, sind statistische Maße, die dazu dienen, eine Datenverteilung in vier gleich große Teile zu unterteilen. Quartile sind nützliche statistische Maße, um die Streuung und Verteilung von Daten zu verstehen, insbesondere wenn es darum geht, Ausreißer oder ungewöhnliche Beobachtungen zu identifizieren, da sie weniger anfällig für Extremwerte sind. Zur Bestimmung der Quartile werden die Beobachtungswerte in aufsteigender Reihenfolge, beginnend mit dem kleinsten Wert, angeordnet. Wenn die Anzahl der Datenpunkte (N) gerade ist, teilt der Median (siehe Median) die Datenreihe in zwei gleiche Hälften. Die Quartile sind dann die Mediane dieser beiden Hälften. Angenommen, wir haben eine kleine Datenreihe des monatlichen Eiscreme-Konsums in Litern: 130,140,150,160,170,130,140,150,160,170. Die Quartile dieser Daten sind $Q_1=140$, $Q_2=150$, $Q_2=150$ (Median), und $Q_3=160$.

R Randomisierung

Unter Randomisierung versteht man die zufällige Zuteilung von Patienten oder Probanden auf verschiedene Gruppen mit dem Ziel, möglichst ähnliche Versuchs- und Kontrollgruppen in Bezug auf personenbezogene Störfaktoren zu erhalten. Dies ist ein entscheidender Schritt in medizinischen Studien, um verallgemeinerbare Aussagen zu ermöglichen und neben der Verblindung eine wichtige Voraussetzung für die wissenschaftliche Validität. Beispielsweise werden die Teilnehmer zufällig entweder der Gruppe mit höheren Temperaturen oder der Kontrollgruppe zugeordnet. Dies soll sicherstellen, dass jegliche beobachteten Effekte auf die unabhängige Variable (Temperaturen) zurückgeführt werden können und nicht auf andere Faktoren.

Reliabilität

Reliabilität, auch Zuverlässigkeit genannt, ist ein wichtiges Kriterium in wissenschaftlichen Untersuchungen. Sie bezieht sich darauf, dass bei wiederholter Durchführung einer Untersuchung unter gleichen Bedingungen konsistente und gleichbleibende Ergebnisse erzielt werden sollten. Reliabilität zeigt an, dass keine zufälligen Fehler oder Schwankungen bei den Messungen auftreten und dass die Ergebnisse zuverlässig reproduzierbar sind. Die Reliabilität in diesem Kontext könnte sicherstellen, dass die erhobenen Daten zum Eiscreme-Konsum und den Temperaturen über mehrere Sommermonate hinweg konsistent sind, um verlässliche Schlussfolgerungen über den Zusammenhang zwischen höheren Temperaturen und dem Eiscreme-Konsum ziehen zu können.

Regressionsanalyse(n)

Die Regressionsanalyse in der Statistik ist ein Verfahren zur Schätzung des Einflusses eines oder mehrerer Merkmale auf eine abhängige Variable unter Konstanthaltung der anderen Einflussgrößen. Die Regressionsanalyse ermöglicht es, die Beziehungen zwischen Variablen zu analysieren und beispielsweise festzustellen, ob eine bestimmte Variable von einer anderen Variable beeinflusst wird und wenn ja, wie stark dieser Einfluss ist. Mithilfe einer Regressionsanalyse lässt sich untersuchen, wie sich der Eiscreme-Konsum quantitativ in Abhängigkeit von den Temperaturen verändert, wodurch eine mathematische Gleichung erstellt wird, die diesen Zusammenhang beschreibt.

Repräsentativ

Der Begriff "repräsentativ" bedeutet, dass die Personen in der Stichprobe in allen Merkmalen den Personen in der Grundgesamtheit entsprechen. Wenn die Stichprobe nicht repräsentativ ist, können die Schlussfolgerungen nur auf den in der Stichprobe vertretenen Personenkreis angewendet werden. Es besteht die Gefahr, dass selektive Stichproben gezogen werden, in denen bestimmte Gruppen über- oder unterrepräsentiert sind, was die Verallgemeinerbarkeit der Ergebnisse einschränkt. Um eine repräsentative Stichprobe zu erhalten, ist es wichtig, die Teilnehmer nach dem Zufallsprinzip aus der Population auszuwählen, da so sichergestellt

wird, dass alle möglichen Merkmale und Besonderheiten, die in der Population vorhanden sind, zu gleichen Anteilen in der Stichprobe enthalten sind. Die Repräsentativität in diesem Zusammenhang bedeutet, dass die ausgewählte Stichprobe von Daten zum Eiscreme-Konsum und den Temperaturen die gesamte Population adäquat widerspiegelt, um generalisierbare Schlussfolgerungen über den Zusammenhang zwischen diesen Variablen ziehen zu können.

S Signifikanz

Signifikanz in der Statistik bezieht sich auf die statistische Signifikanz eines in einer Stichprobe gefundenen Effekts, den man auf die gesamte Population verallgemeinern möchte. Signifikante Ergebnisse gelten als systematisch und nicht als zufällig, wobei die Signifikanz keine Aussage über die inhaltliche Bedeutung des Effekts macht. In Bezug auf Eiscreme-Konsum und Temperaturen würde Signifikanz darauf hinweisen, dass die beobachteten Unterschiede oder Zusammenhänge in den Daten nicht zufällig sind. Eine signifikante Beziehung zwischen höheren Temperaturen und einem erhöhten Eiscreme-Konsum würde darauf hindeuten, dass diese Verbindung nicht durch reinen Zufall erklärt werden kann und somit relevant ist.

Signifikanzniveau

Das Signifikanzniveau, oft als α bezeichnet, ist ein Schwellenwert in der Statistik, der die Wahrscheinlichkeit angibt, mit der ein Ergebnis als statistisch signifikant angesehen wird. Es wird typischerweise auf einen Wert wie $\alpha = 0,05$ oder $\alpha = 0,01$ festgelegt. Dies bedeutet, dass ein Ergebnis als signifikant angesehen wird, wenn die Wahrscheinlichkeit, es unter Annahme der Nullhypothese zu erhalten, diesen Schwellenwert nicht überschreitet. Das Signifikanzniveau ist ein Maß für die Sicherheit, mit der ein Ergebnis als nicht zufällig angesehen wird, und hat direkten Einfluss auf die Verwerfung der Nullhypothese und die Annahme der Alternativhypothese bei statistischen Tests.

Skalierung

Die Merkmale können nominal (siehe nominal), ordinal (siehe ordinal) oder metrisch (siehe metrisch) skaliert sein.

Standardabweichung

Ein Maß für die Streuung oder Variabilität von Datenpunkten in einer Datenmenge. Die Standardabweichung gibt an, wie weit die einzelnen Werte eines Merkmals im Durchschnitt von ihrem Mittelwert (siehe Mittelwert) entfernt sind. Mit anderen Worten, die Standardabweichung zeigt, wie die gemessenen Werte um den Mittelwert verteilt sind. Die Standardabweichung wird aus der Quadratwurzel der Varianz (siehe Varianz) berechnet. Die Standardabweichung des Eiscreme-Konsums könnte beispielsweise 3 Liter betragen, was darauf hinweist, wie stark die einzelnen Monatswerte um den Durchschnitt schwanken.

Statistischer Test

Soll bei einer wissenschaftlichen Fragestellung z.B. ein Vergleich zwischen zwei oder mehreren Gruppen untersucht werden, kann ein statistischer Test durchgeführt werden. Er berechnet die Wahrscheinlichkeit, mit der die beobachteten Daten auftreten, wenn die Nullhypothese wahr ist. Mit Hilfe von angenommenen Datenverteilungen wird dieser Wert mit den erwarteten Verteilungen verglichen, um zu entscheiden, ob die Beobachtungen zufällig oder statistisch signifikant sind.

Statistik

Statistik ist die systematische Analyse von Zahlen und Daten, um in der Forschung Schlussfolgerungen zu ziehen und Fragen zu beantworten. Dabei werden häufig Hypothesen verwendet, um Daten gezielt zu untersuchen und Erkenntnisse zu gewinnen.

Stichprobe

Eine Stichprobe ist eine ausgewählte Untergruppe der Grundgesamtheit (siehe Population), die tatsächlich in einer Studie untersucht wird. Aus praktischen Gründen ist es oft nicht möglich, die gesamte Population zu untersuchen. Daher wählen die Forschenden eine repräsentative (siehe repräsentativ) Gruppe aus, die als Stichprobe bezeichnet wird. Die Ergebnisse der Stichprobe werden verwendet, um Schlussfolgerungen über die gesamte Population zu ziehen (siehe Generalisierung). Wenn du Daten von 100 zufällig ausgewählten Einwohnern sammelst, repräsentiert diese Stichprobe die Gesamtbevölkerung der Stadt.

T t-Test

Der t-Test ist ein statistisches Werkzeug, das zum Vergleich der Mittelwerte von einer oder zwei Gruppen oder Populationen mit Hilfe von Hypothesentests verwendet wird. Angenommen, es wird überprüft, ob es einen signifikanten Unterschied im Eiscreme-Konsum zwischen zwei Gruppen von Menschen gibt: einer Gruppe, die höheren Temperaturen ausgesetzt war, und einer anderen Gruppe, die niedrigeren Temperaturen ausgesetzt war. Ein t-Test könnte durchgeführt werden, um festzustellen, ob die durchschnittlichen Eiscreme-Konsumwerte zwischen den beiden Gruppen statistisch signifikant voneinander abweichen.

V Validität

Die Validität ist ein wichtiges Kriterium in wissenschaftlichen Untersuchungen und bezieht sich auf die Genauigkeit und Richtigkeit der Messungen und darauf, ob die Daten die untersuchten Fragen tatsächlich angemessen beschreiben. Wenn eine Untersuchung valide ist, bedeutet dies, dass sie tatsächlich das gemessen hat, was gemessen werden sollte, und dass die Ergebnisse zuverlässig und zutreffend sind. Die Validität unterscheidet sich von der Reliabilität, da selbst sehr genaue Messungen nicht unbedingt valide sind, wenn sie die falschen

Aspekte erfassen. Zum Beispiel wäre eine Messung valide, wenn sie tatsächlich den tatsächlichen Eiscreme-Konsum in verschiedenen Temperaturen genau repräsentiert und nicht von anderen Faktoren beeinflusst wird.

Variable

Eine Variable repräsentiert eine Gruppe beobachtbarer oder messbarer Merkmalsausprägungen (siehe Merkmalsausprägungen), die unterschiedliche Werte annehmen. Es kann zwischen abhängigen Variablen und unabhängigen Variablen unterschieden werden. Die Ausprägung der abhängigen Variablen, die im Forschungsprozess erklärt werden soll, wird durch die Ausprägung der unabhängigen Variablen bestimmt (unabhängige Variable → abhängige Variable). (Unabhängige Variable (Temperaturen) und die abhängige Variable (Eiscreme-Konsum))

Varianz

Die Varianz ist ein Maß dafür, wie die Werte einer Gruppe um den Mittelwert verteilt sind. Sie wird berechnet, indem die Abweichungen der einzelnen Werte vom Mittelwert quadriert, die quadrierten Abweichungen addiert und durch die Anzahl der Werte in der Gruppe dividiert werden. Die Varianz ist das Quadrat der Standardabweichung (siehe Standardabweichung). Die Varianz in der Statistik hilft zu verstehen, wie sich die Datenpunkte in einer Gruppe voneinander unterscheiden. Wenn die Varianz groß ist, gibt es große Unterschiede zwischen den Datenpunkten. Wenn die Varianz klein ist, sind die Zahlen in der Gruppe ziemlich ähnlich. Die Varianz ist also ein Maß für die Verteilung der Zahlen in einer Gruppe. Angenommen, in einer Untersuchung zum Eiscreme-Konsum analysiert man die Varianz der monatlichen Daten in Bezug auf die Temperaturen. Die Varianz gibt Aufschluss darüber, wie stark die einzelnen Werte um den Durchschnitt schwanken. Eine höhere Varianz deutet auf signifikante Schwankungen im Eiscreme-Konsum hin, während eine niedrigere Varianz auf eine gleichmäßigere Verteilung der Werte hinweisen würde.

Z z-Wert

Der “Z-Wert” in der Statistik misst die Abweichung eines Datenpunktes vom Mittelwert einer Verteilung. Er gibt an, ob dieser Datenpunkt für die Verteilung typisch ist oder von den anderen Datenpunkten abweicht. Ein Z-Wert von 0 bedeutet, dass der Datenpunkt dem Durchschnitt entspricht. Ein positiver Z-Wert zeigt an, dass der Datenpunkt über dem Durchschnitt liegt, während ein negativer Z-Wert anzeigt, dass der Datenpunkt unter dem Durchschnitt liegt. Angenommen, in einer Analyse des Eiscreme-Konsums in Bezug auf die Temperaturen wird der z-Wert verwendet, um die Werte zu standardisieren und vergleichbar zu machen. Der z-Wert gibt an, wie viele Standardabweichungen ein einzelner Datenpunkt vom Durchschnitt entfernt ist. Ein positiver z-Wert zeigt an, dass der Eiscreme-Konsum über dem Durchschnitt liegt, während ein negativer Wert darauf hinweist, dass er darunter liegt. Der z-Wert ermöglicht eine standardisierte Darstellung der Daten und erleichtert den Vergleich von Werten in unterschiedlichen Maßeinheiten.