Grundlagen_der_Visualisierung_czlabinger

Zlabinger Christof

2023-11-13

Grundlagen der Visualisierung

Merkmale und Daten

Qualitativ

Alles was nur in Kategorien eingeteilt werden kann. Es gibt ein Maximum wie viele es gibt. Es kann keine "Zwichen Kategorien geben" z.B.: 1, 1.5, 2

nominal Keine Vordefinierte Anordnung. z.B.: Geschlecht, Farben

ordinal Klare Ordnung z.B.: Noten Es kann nicht mir ihnen gerechnet werden. Selbe differenzen koennen verschieden gewichtet sein.

Quantitativ

Koennen gemessen und verglichen werden. Es kann mit ihnen gerechnet werden.

intervallskalierte Differenzen koennen ausgerechnet werden aber keine vielfachheiten interpretieren kann. z.B.: Das Jahr 2000 ist spaeter als das Jahr 1000 aber das Jahr 1000 ist nicht das doppelte vom Jahr 500. Mit dem Jahr -500 kann keine sinnvolle Berechnung durchgefuehrt werden.

Man kann nicht sagen -5°C ist das -1 fache von 5°C. Wenn die Werte 0 oder negative Zahlen beinhalten koennen kann nicht von vielfachen gesprochen werden sondern nur von differenzen. 5°C ist um 10°C mehr als -5°C

rationalskaliert Daten die nur positiv sein koennen. z.B.: Kelvin da es einen absoluten 0-punkt hat der in der Natur nicht erreicht werden kann. 1000K sind das doppelte von 500K

diskret z.B.: Zaehler Variabeln: 0,1,2,...

stetig Belibige Werte innerhalb eines Intervalls, eingeschraenkt durch die Messgenauigkeit.

Datenmatrix

In den Spalten stehen die Merkmale aka Variabeln In den Zeile stehen die UNtersuchungseinheiten

Schaetzung und Darstellung von qualitativer Messungen

absolute Haeufigkeit

relative Haeufigkeit = absolute/gesammt

Koennen als Balken- oder Tortendiagramm darkestellt werden

Nichkummulierte Darstellung so wie sie sind gegenueber

Laengen koennen gut geschaetzt werden. Winkel nicht. Vergleich von % kein Tortendiagramm Balekendiagramm besser geeignet. Tortendiagramm gut um zu sehen wann es eine mehrheit gibt und um wie viel.

Kumulative Addieren alle Gruppen bis zu einer Bestimmten auf. Wennes eine zu grunde liegende Anordnung gibt. Mehr sinn fuer ordinal als nominal.

Eigenschaften von quantitativen Daten:

Lage - Modus Unimodalität ist die Voraussetzung für alle anderen Lageschätzer wie den Mittelwert. Bimodalität oder Multimodalität müssen die Daten in verschiedene "Lager" aufgeteilt werden, um sie besser zu beschreiben.

Streuung - Varianz & Standardabweichung Varianz: Sigmaˆ2 = 1/n*sum(i, n, (datenpunkkt - mittelwert)ˆ2)(nicht stabil)

Standardabweichung: Sigma = sqrt(Sigmaˆ2) (nicht stabil)

Interquartildistanz: Distanz zwischen dem 25%-Quantil (1. Quartil) und dem 75%-Quantil (3. Quartil) (stabil)

Spannweite: Maximalwert - Minimalwert (nicht stabil)

Symmetrie und Schiefe

- Normalverteilung: Symmetrisch mit einem Gipfel.
- Gleichverteilung: Symmetrisch ohne Peak - alle Werte sind gleich hoch.
- Bimodal symmetrisch: Zwei Peaks; Mittelwert und Median sind nicht sinnvoll, wenn sie bei 0 sind.
- Linksschief: Die Daten steigen rechts im Graphen stark an und fallen links langsam ab.
- Rechtsschief: Die Daten steigen links im Graphen stark an und fallen rechts langsam ab.
- Gewicht an den Rändern (Kurtosis)
  - Gibt an, wie stark Daten von ihrem Zentrum abweichen; erkennbar am Verschieben des Medians bei linksschiefen oder rechtsschiefen Daten.

Quantilschätzer

Gibt einen Zahlenwert für einen beliebigen Wert sigma sigma zwischen 0 und 1 aus, der dem Anteil aller Messwerte im Datensatz entspricht. Der Median ist beispielsweise das 50%-Quantil. Das Minimum (0%-Quantil) und Maximum (100%-Quantil) sind ebenfalls enthalten. Quantilschätzer geben an, wie viele Daten unbrauchbar sein können, aber dennoch funktionieren. Zum Beispiel können beim Median 50% der Daten unbrauchbar sein, und das Ergebnis ist dennoch sinnvoll. Der Anteil wird durch sigma sigma angegeben.

Dichtefunktion

Eine Dichtefunktion ist eine Funktion, bei der die Summe der n-ten Potenz der Datenwerte die Fläche unter dem Graphen ist.

2. Video

Arithmetisches Mittel

Alle Werte aufaddieren und durch die Gessamtanzahl dividiert.

`mean()`

Median

Der Wert in der Mitte bei Geodnetern Daten.

`median()`

Modus

Wert mit der groessten Haeufigkeit

`mode()`

Schaezung und Darstellung von quantitativer Messungen

Zaehlvariablen

Nur in ganzen Messeinheiten Zaehlbar. Wie oft kommt welche Grundanyahl vor. Anhand dessen absolute und relative haeufigkeit. Auch aufsummierbar.

Stetige nummerische Variablen

Nicht wahrscheinlich dass 2 mal der gleiche Wert vorkommt. Somit keine Kategorien bilden. z.B: Messung von Dichten. Zusammenfassung in Teilbereiche. (Intervalle) Mit Intervallen wieder als Balkendiagramm darstellbar.(Histogramm um zu sehen ob Daten Unimodal oder Multimodal sind)

Kumulative Verteilungsfunktion Hilft moegliche Aussreiser zu erkennen wenn sie weit Weg von den anderen Werten sind.

Boxplot Zusammenfassung der Quartile. Box laeuft vom 1. bis zum 3. Quantil. (25%-75%) In der Box ist der Median 50% Quantil. Robust gegenueber aussreisern. Ausserhalb der Box sind die "Whiskers". Interquatilsdistanz ist die Distanz von 1. Quantil bis zum 3. Quantil. Laufen bis zur 1.5 fachen interquatilsdistanz. Schneidet aber beim lezten Wert innerhab von 1.5 Box laengen ab.

Normal Q-Q Plot

Vergleicht Messwerte mit den Theoretischen Quantilen Wie kann ich feststellen ob es eine Normalverteilung gibt. Normalverteilt wenn die Punkte auf der Referenz liegen (Mit gewisser Tolleranz). Wenn sehr viele Punkte gibt die nicht auf der Referenz liegen sind es keine Normalverteilte Daten.

Ca um den Median zentriert. Leichte Kannte.

Bei schwerer Kante Boxplot nicht mehr gut um aussreiser zu erkennen.