**Projekt I**

**Statistik**

**Fachkonzept I**

**Beschreibung der Algorithmen**

Inhaltsverzeichnis

[1. Fallbeispiel 3](#_Toc473134550)

[2. Absolute und relative Klassenhäufigkeit 3](#_Toc473134551)

[3. Klassenmitte 4](#_Toc473134552)

[4. Histogramm 4](#_Toc473134553)

[4.1 Klassenbreite 4](#_Toc473134554)

[5. Empirische Verteilungsfunktion 5](#_Toc473134555)

[6. Mittelwert 6](#_Toc473134556)

[7. Median 6](#_Toc473134557)

[8. Quantil 7](#_Toc473134558)

[9. Mittlere absolute Abweichung zu einem Wert z 8](#_Toc473134559)

[10. Varianz 8](#_Toc473134560)

[10.1 Standardabweichung: 8](#_Toc473134561)

[11. Gini-Koeffizient 9](#_Toc473134562)

# Fallbeispiel

Das hier beschriebene Fallbeispiel, wird in allen folgenden Beschreibungen der Algorithmen verwendet:

In einer Schulklasse werden die Sockengrößen der einzelnen Schüler gemessen und notiert. Diese haben jedoch keine feste Größe, sondern nur Bereichsgrößen.

Die Stichprobe *x* setzt sich dabei aus mehreren Bereichsgrößen, sog. Intervallen zusammen. Diese nennt man Klassen :

Der Stichprobenumfang beträgt mit der Klassenanzahl :

# Absolute und relative Klassenhäufigkeit

Die absolute Klassenhäufigkeit ergibt sich aus der Anzahl der einzelnen Beobachtungswerte, welche in die jeweilige Klasse der Stichprobe *x* fallen:

Die relative Klassenhäufigkeit bildet sich aus der Division der absoluten Häufigkeit durch den Stichprobenumfang *n*:

Allgemein gilt also:

# Klassenmitte

Die Klassenmitte lässt sich wie folgt berechnen, wenn das Teilintervall als beschrieben wird:

Somit sind die Klassenmitten des Beispiels:

# Histogramm

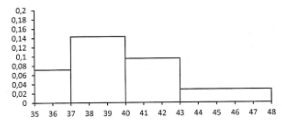
Ein Histogramm beschreibt die Intervalle der Klassen einer Stichprobe in Rechtecken, deren Flächeninhalt proportional zu den entsprechenden Häufigkeiten ist.

Die Höhe des Rechtecks der Klasse ist dann , wobei die Klassenbreite bezeichnet. Beispiel:

## 4.1 **Klassenbreite**

Beispiel:

Damit kann das Diagramm wie folgt gezeichnet werden:

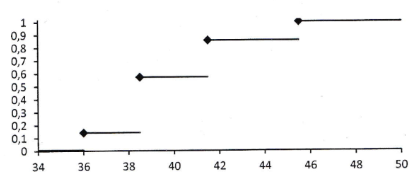


# Empirische Verteilungsfunktion

Für das Zeichnen der empirischen Verteilungsfunktion werden die Klassenmitten sowie die relativen Klassenhäufigkeiten benötigt:

Beispiel:

Das folgende Schaubild zeigt diese Funktion:



Dabei stellen die Intervallgrenzen die Länge des Strichs parallel zur x-Achse dar und die E(x) die entsprechenden Höhen der Linie.

# Mittelwert

Der Mittelwert lässt sich für Stichproben ohne Klasseneinteilung wie folgt berechnen:

Für die Berechnung des Mittelwerts von Stichproben mit Klasseneinteilung werden jeweils die Klassenmitten zur Berechnung verwendet.

Der Mittelwert kann also auf zwei verschiedene Weisen berechnet werden:

Oder

# Median

Der Median liegt in der Mitte einer geordneten Stichprobe. Also mindestens 50% der Werte sind kleiner oder gleich dem Median und gleichzeitig mindestens 50% der Werte größer oder gleich dem Median.

Um den Median bestimmen zu können werden wieder die Klassenmitten bestimmt:

Außerdem sind die relativen Häufigkeiten bereits bestimmt worden:

Bestimmung des Medians anhand des Fallbeispiels:

Schlussfolgernd fällt der Median in die Klasse

Allgemein gilt:

Der Median lässt sich durch addieren der relativen Klassenhäufigkeiten ermitteln. Wird der Wert 0,5 das erste Mal überschritten, fällt der Median auf die entsprechende Klasse.

Die Klasse wird nun näher betrachtet, um den Median wie folgt zu bestimmen:

Also

Der Median der Stichprobe x beträgt also 39,5.

# Quantil

Das Quantil lässt sich allgemein wie folgt berechnen:

Bei einem Vergleich mit der Berechnung des Medians, fällt auf, dass die Berechnung für das Quantil analog erfolgt.

Im Folgenden wird die Rechnung anhand des Beispiels erläutert, dabei sei = also wird das Quantil gesucht:

Nun muss die Bedingung erfüllt sein.

Somit fällt das Quantil in . Es folgt anschließend die Berechnung mit der allgemeinen Formel analog zur Bestimmung des Median:

Dementsprechend sind die Quantile:

# Mittlere absolute Abweichung zu einem Wert z

Geeignete Werte für sind Median und Mittelwert, aber auch die beiden Randwerte.

Im Folgenden wird die Mittlere absolute Abweichung zum Wert berechnet:

# Varianz

Allgemein lässt sich die Varianz wie folgt berechnen:

Die Varianz wird im Folgenden anhand des Fallbeispiels berechnet:

## 10.1 Standardabweichung:

Das zu untersuchende Merkmal sei metrisch skaliert. Dann bildet sich die Standardabweichung wie folgt:

Da folgt für die Beispielrechnung:

# Gini-Koeffizient

1. Klassenmitte bestimmen

Gegeben: Kn=( tX, tX+1], …

KlassenmitteKn =

1. Klassenmitten der Größe nach ordnen und Reihenfolge festlegen
2. Tabelle ausfüllen

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| m | um | vm | um – um-1 | vm + vm-1 | (um – um-1)\*( vm + vm-1) |
| 0 | 0 | 0 | 0 | 0 | 0 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

! Die erste Zeile der Tabelle beinhaltet immer ausschließlich „0“er!

m = geordnete Reihenfolge der Klassen

um= relativer Anteil der Klasse m an n Klassen 🡪

vm= relativer Anteil der Summe von Klasse m an der Gesamtsumme aller Klassen 🡪

1. **Gini Koeffizient berechnen**

Beispiel:

1. Klassenmitte bestimmen
2. Klassenmitten der Größe nach ordnen und Reihenfolge festlegen
3. Tabelle ausfüllen

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| m | um | vm | um – um-1 | vm + vm-1 | (um – um-1)\*( vm + vm-1) |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |

1. Gini Koeffizient berechnen

Der Gini Koeffizient beträgt somit .