

Beschreibende Statistik I: Statistische Begriff und Kennzahlen

In der beschreibenden Statistik werden sehr oft Fachbegriffe verwendet. Das Wissen über dieses Vokabular ist der Schlüssel zum Verständnis dieses Themengebietes. Im Folgenden werden mehrere statistische Kennzahlen und Grundbegriffe der beschreibenden Statistik erklärt und anhand eines Beispiels erklärt.

1) Urliste

Die Urliste ist das Endergebnis der Erhebung der Daten. Sie ist generell ungeordnet.

5	7	2	9	4	7	3	7
---	---	---	---	---	---	---	---

2) Geordnete Liste

Von einer geordneten Liste spricht man, wenn die Daten aus der Urliste aufsteigend in Reihung gebracht werden, sofern sie in Reihung zu bringen sind.

2	3	4	5	7	7	7	9
---	---	---	---	---	---	---	---

3) Arithmetisches Mittel

Das arithmetische Mittel \bar{x} ist ein statistisches Zentralmaß, das umgangssprachlich den Durchschnitt oder den Mittelwert meint. Mathematisch gesehen ist das arithmetische Mittel die Summe der Werte durch die Anzahl der Werte.

Anzahl der Werte = $n \rightarrow$ hier $n = 8$

$$\bar{x} = \frac{\text{Summe der Werte}}{\text{Anzahl der Werte}} = \frac{5+7+2+9+4+7+3+7}{8} = \frac{44}{8} = 5,5$$

\rightarrow Das arithmetische Mittel lautet also 5,5.

4) (Empirische) Varianz

Die empirische Varianz ist ein statistisches Streuungsmaß und meint die mittlere, quadratische Abweichung aller Daten vom arithmetischen Mittel. Hiermit kann abgeschätzt werden, wie weit die Daten der Liste vom Mittelwert streuen. x_1, x_2, x_3, \dots sind die Datenelemente einer Liste.

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 \dots (x_n - \bar{x})^2}{n}$$

$$s^2 = \frac{(2-5,5)^2 + (3-5,5)^2 + (4-5,5)^2 + (5-5,5)^2 + (7-5,5)^2 + (7-5,5)^2 + (7-5,5)^2 + (7-5,5)^2 + (9-5,5)^2}{8}$$

$$s^2 = \frac{40}{8} = 5$$

\rightarrow Die empirische Varianz lautet also 5.

5) (Empirische) Standardabweichung

Die empirische Standardabweichung ist ein statistisches Streuungsmaß und gibt den mittleren Abstand der Werte zum arithmetischen Mittel an. Sie ist die Wurzel der empirischen Varianz.

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 \dots (x_n - \bar{x})^2}{n}}$$

$$s = \sqrt{\frac{(2-5,5)^2 + (3-5,5)^2 + (4-5,5)^2 + (5-5,5)^2 + (7-5,5)^2 + (7-5,5)^2 + (7-5,5)^2 + (7-5,5)^2 + (9-5,5)^2}{8}}$$

$$s = \sqrt{\frac{40}{8}} = \sqrt{5} = 2,236$$

→ Die empirische Standardabweichung lautet also gerundet 2,2.

6) Modus oder Modalwert

Der Modus ist ein Zentralmaß und gibt den Wert an, der in der Liste am häufigsten vorkommt.

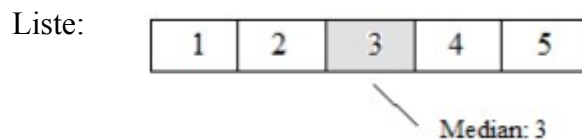
In unserem Beispiel kommt der Wert 7 insgesamt dreimal vor.

→ Der Modus lautet also 7.

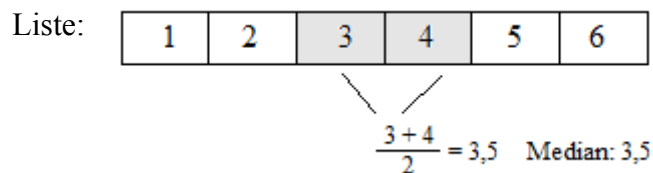
7) Median oder Zentralwert

Der Median m ist ein Zentralmaß und gibt den mittleren Wert der geordneten Liste an.

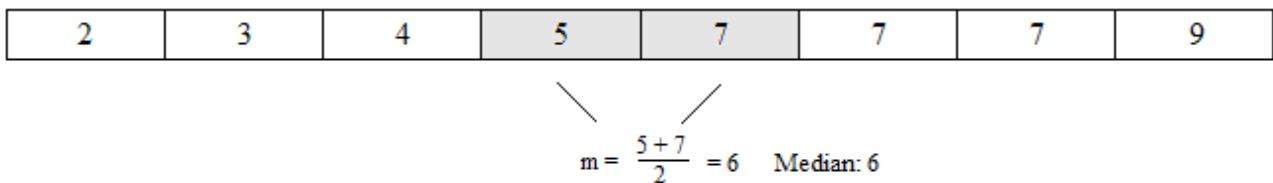
- besteht die Liste aus einer ungeraden Anzahl von Werten, so ist der Median eindeutig, wie hier.



- besteht die Liste aus einer geraden Anzahl von Werten, so ist der Median das arithmetischen Mittel der beiden mittleren Werte, wie hier.



Zurück zu unserer Datenliste.



→ Der Median lautet also 6.

8) Minimum und Maximum

Das Minimum \min ist der kleinste Wert der Liste. Das Maximum \max ist der größte Wert der Liste.

→ Das Minimum ist also 2.

→ Das Maximum ist also 9.

9) Spannweite

Die Spannweite Sp ist ein statistisches Streuungsmaß und gibt an in welchem Wertebereich die Daten der Liste liegen. Es wird als Differenz von Maximum und Minimum errechnet.

$$Sp = \max - \min = 9 - 2 = 7$$

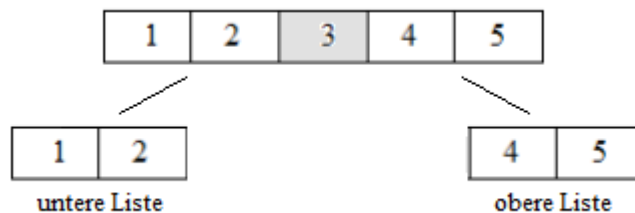
→ Die Spannweite lautet als 7.

10) Untere Liste und obere Liste

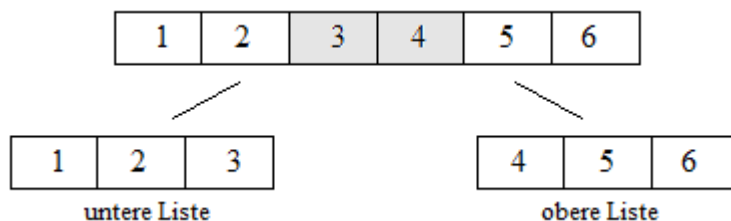
Der Median teilt die geordnete Liste in zwei Teile: in eine untere und eine obere Liste.

Konventionell gibt es hierbei strikte Vorgehensweisen.

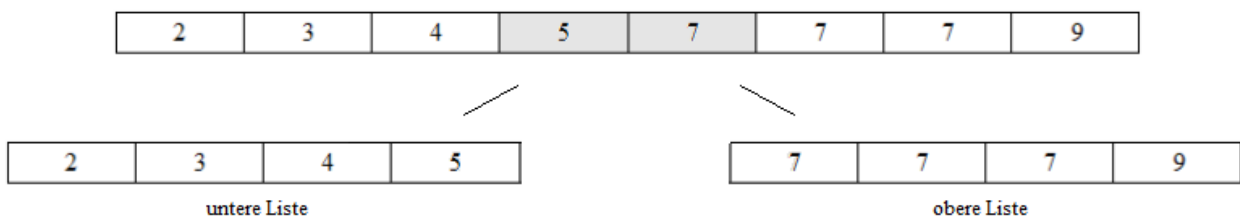
- besteht die Liste aus einer ungeraden Anzahl von Werten, so gehört der Median weder zur unteren noch zur oberen Liste.



- besteht die Liste aus einer geraden Anzahl von Werten, so errechnet sich der Median aus den beiden mittleren Werten. Der kleinere der beiden Medianwerte kommt zur unteren Liste, der größere zur oberen Liste.



Zurück zu unserer Datenliste.



→ Die untere Liste lautet also 2, 3, 4, 5

→ Die obere Liste lautet als 7, 7, 7, 9

11) Quartile

Die Quartile teilen die geordnete Liste in vier Teile mit gleich vielen Daten.

Man spricht von einem unteren Quartil (Q_1 oder 25%-Quartil), dem Median (Q_2 oder 50%-Quartil) und dem oberen Quartil (Q_3 oder 75%-Quartil).

Das untere Quartil ist der Median der unteren Liste, das obere Quartil ist der Median der oberen Liste. Die Regeln für die Bildung des Median und die Einteilung in die untere und obere Liste sind unbedingt zu beachten.

untere Liste				obere Liste				
	Q_1		$m = Q_2$		Q_3			
2	3	4	5	7	7	7	9	
	3,5		6		7			

→ Das untere Quartil (Q_1) lautet also 3,5.

→ Das obere Quartil (Q_3) lautet also 7.

12) Interquartilspannweite oder Interquartilsabstand

Die Interquartilspannweite I_{sp} ist ein statistisches Streuungsmaß, das den Abstand zwischen oberen und unteren Quartil angibt. Dadurch erkennt man in welchem Bereich die mittleren 50% der Werte der Datenliste streuen. Errechnet wird die Interquartilspannweite mittels Differenz von Q_3 und Q_1 .

$$I_{sp} = Q_3 - Q_1 = 7 - 3,5 = 3,5$$

→ Die Interquartilspannweite lautet also 3,5.

13) Fünf-Daten-Zusammenfassung

Für die spätere Erstellung eines Boxplots werden fünf Daten benötigt. Das Minimum (min), das untere Quartil (Q_1), der Median (m), das obere Quartil (Q_3) und das Maximum (max).

Sie sollte übersichtlich in einer Tabelle angeschrieben werden.

min	2
Q_1	3,5
m	6
Q_3	7
max	9

14) Zentralmaße

Zentralmaße sind Werte, die repräsentativ für eine Liste stehen. Sie können in der Liste vorkommen oder auch nicht. Unter dem Begriff Zentralmaße fallen das arithmetische Mittel (\bar{x}), der Modus (mod) und der Median (m).

Um die Übersicht zu wahren wird in einer Tabelle dargestellt.

\bar{x}	5,5
mod	7
m	6

15) Streuungsmaße

Streuungsmaße sind Werte, die angeben wie weit Werte einer Liste verteilt sind bzw. wie weit sie um Mittelwert streuen. Streuungsmaße sind die empirische Varianz (s^2), die empirische Standardabweichung (s), die Spannweite (Sp) und die Interquartils Spannweite (I_{sp}).

Auch hier wird übersichtlich in einer Tabelle aufgelistet.

s^2	5
s	2,2
Sp	7
I_{sp}	3,5

Übung 1: Gegeben sind zwei Datensätze. Arbeite die Punkte 1) bis 15) der Anleitung für die gegebenen Datensätze durch.

a) Bei einem Vokabeltest im Fach Italienisch werden 12 Vokabeln abgefragt. Das Ergebnis der einzelnen Schüler werden in folgender Urliste dargestellt.

7	8	8	12	6	8	7	6	10	5
---	---	---	----	---	---	---	---	----	---

b) Bei Verkehrskontrollen wird in 11 Gemeinden kontrolliert, wie viele Insassen nicht angeschnallt sind. Das Endergebnis der Erhebung wird für alle Gemeinden in folgende Urliste eingetragen.

4	5	8	17	4	13	8	4	6	10	9
---	---	---	----	---	----	---	---	---	----	---

Übung 2: Gegeben ist nun eine Fünf-Daten-Zusammenfassung.

min	Q_1	m	Q_3	max
3	5	7	12	18

- Erstelle einen Datensatz aus 12 Werten, die zur gegebenen Fünf-Daten-Zusammenfassung passt.
- Erstelle einen Datensatz aus 13 Werten, die zur gegebenen Fünf-Daten-Zusammenfassung passt.
- Erstelle einen Datensatz aus 15 Werten, die zur gegebenen Fünf-Daten-Zusammenfassung passt.