Beschreibende Statistik II: Häufigkeiten und Diagramme

Urliste

15	8	11	11	13	9	3	8	9	1	5	12	6	9	8	2	11	9	10	12	2	2	3

16) Klassen

Die Einteilung der Daten in Klassen kann vorgegeben oder selbst zu wählen sein. Dies geschieht oft willkürlich, wobei man schon auf die Sinnhaftigkeit der Klassen achten sollte. Jeder Wert sollte genau in eine Klasse gehören, deshalb darf es keine Inervallüberschneidungen geben.

Hier ist die Klasseneinteilung vorgegeben. Angegeben werden die Klassen meist in Intervallschreibweise.

Klasseneinteilung: 1. Klasse: [0 ; 3], 2. Klasse: (3 ; 7], 3. Klasse: (7 ; 11], 4. Klasse: (11 ; 15]

17) Strichliste

Zur Zählung der Daten jeder Klasse wird eine Strichliste angelegt. Dabei wird für jeden Wert ein Strich in der jeweiligen Klasse gesetzt. Zusätzlich werden die Striche in Fünferbündel zusammengefasst, um die Zählung zu erleichtern.

18) Absolute Häufigkeit

Die absolute Häufigkeit ist die Anzahl der Werte jeder Klasse, also die Zahl, die zustandekommt, wenn die Striche gezählt werden.

19) Relative Häufigkeiten

Die relative Häufigkeit ist der Quotient aus der absoluten Häufigkeit einer Klasse und der Summe aller absoluten Häufigkeiten (bzw. dem Stichprobenumfang n).

20) Prozentuelle Häufigkeit

Die prozentuelle Häufigkeit ergibt sich aus dem Produkt des Wertes der relativen Häufigkeit jeder Klasse mit 100, relative Häufigkeit mal 100.

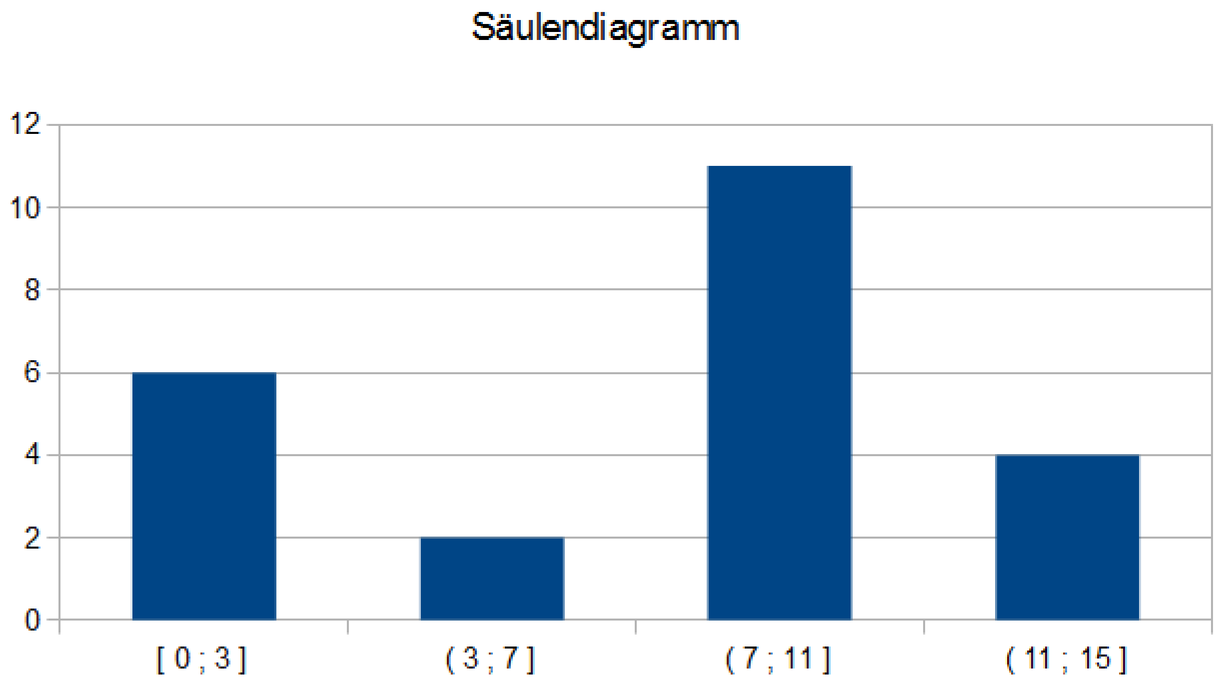
21) Häufigkeitstabelle

Aus praktischen Gründen werden die zuvor erwähnten Werte in einer Tabelle zusammengefasst.

	Strichliste	Absolute Häufigkeit	Relative Häufigkeit	Prozentuelle Häufigkeit
[0 ; 3]	卌 I	6	6/23 = 0,261	26,1%
(3 ; 7]	II	2	2/23 = 0,087	8,7%
(7 ; 11]	卌 卌 I	11	11/23 = 0,478	47,8%
(11 ; 15]	IIII	4	4/23 = 0,174	17,4%
Summe		23	23/23 = 1	100%

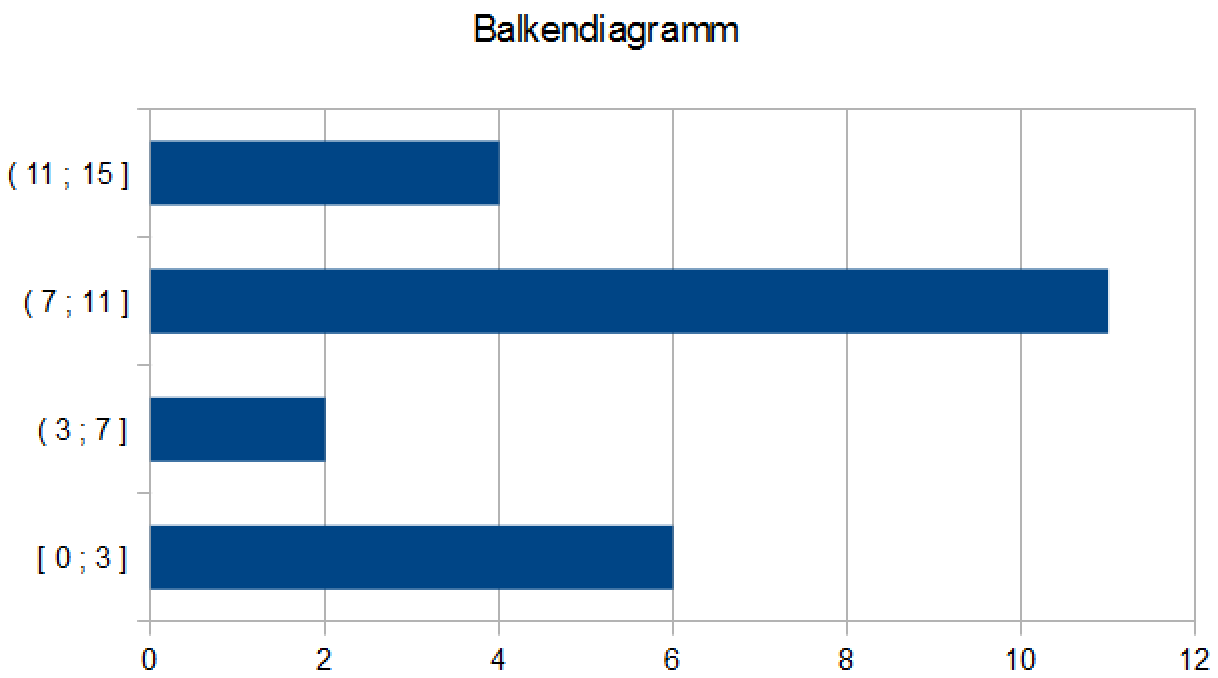
22) Säulendiagramm oder Stabdiagramm

Bei dieser Art von Diagramm werden die absoluten Häufigkeiten mittels (vertikalen Säulen) graphisch dargestellt. Zu achten ist dabei auf eine vorteilhafte Skalierung der y-Achse.



23) Balkendiagramm

Bei dieser Art von Diagramm werden die absoluten Häufigkeiten mittels (horizontalen Säulen) graphisch dargestellt. Zu achten ist dabei auf eine vorteilhafte Skalierung der x-Achse.

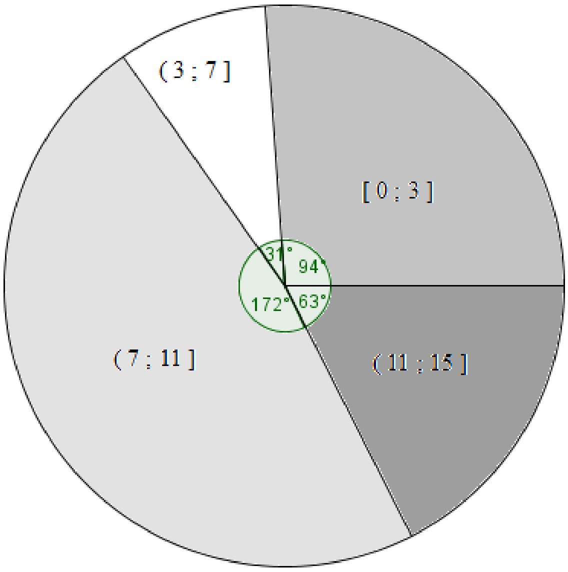


24) Kreisdiagramm

Bei dieser Art von Darstellung werden die prozentuellen Häufigkeiten mittels Kreissektoren in einem 360°-Kreis dargestellt. Dazu braucht man das Maß des Zentriwinkels aller Kreissektoren.

$$1\% = 3{,}6°$$

	Prozentuelle Häufigkeit	Maß des Zentriwinkels
[0 ; 3]	26,1%	**≈ 94°**
(3 ; 7]	8,7%	**≈ 31°**
(7 ; 11]	47,8%	**≈ 172°**
(11 ; 15]	17,4%	**≈ 63°**
Summe	100%	360°



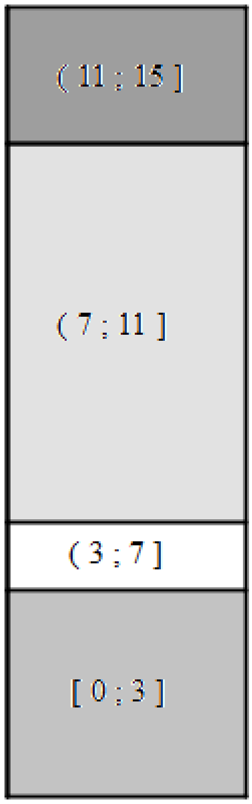
25) Prozentsteifendiagramm

Bei dieser Art von Darstellung werden die prozentuellen Häufigkeiten mit einem 100 mm hohen Rechteck dargestellt. Dazu braucht man die Höhe der einzelnen Abschnitte.

Jedes Prozent entspricht genau einem Millimeter Rechteckhöhe.

$$1\% = 1 \text{ mm}$$

	Prozentuelle Häufigkeit	Höhe des Streifens
[0 ; 3]	26,1%	**26,1 mm**
(3 ; 7]	8,7%	**8,7 mm**
(7 ; 11]	47,8%	**47,8 mm**
(11 ; 15]	17,4%	**17,4 mm**
Summe	100%	100m



26) Kastenschaubild (Boxplot)

Das Kastenschaubild ist eine graphische Veranschaulichung der Fünf-Daten-Zusammenfassung. Es ermöglicht einen übersichtlichen Einblick auf die Verteilung der Daten.

Urliste

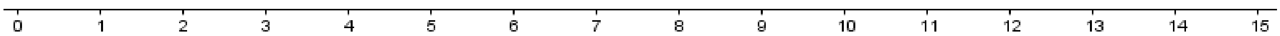
15	8	11	11	13	9	3	8	9	1	5	12	6	9	8	2	11	9	10	12	2	2	3

Geordnete Liste

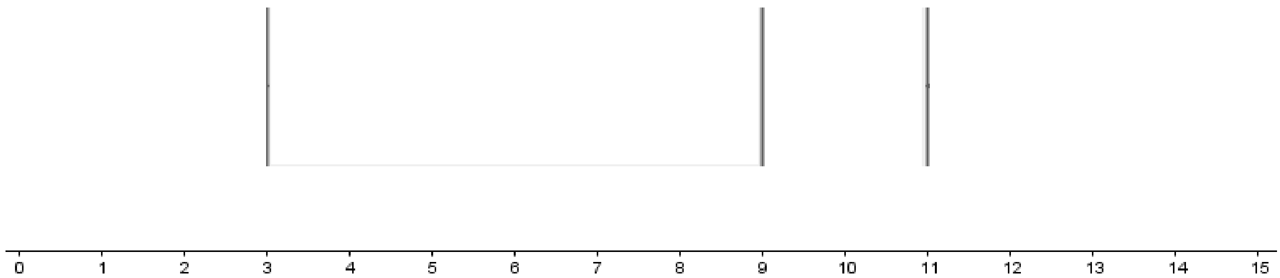
1	2	2	2	3	3	5	6	8	8	8	9	9	9	9	10	11	11	11	12	12	13	15
min					Q1						m						Q3					max

min	Q_1	m	Q_3	max
1	3	9	11	15

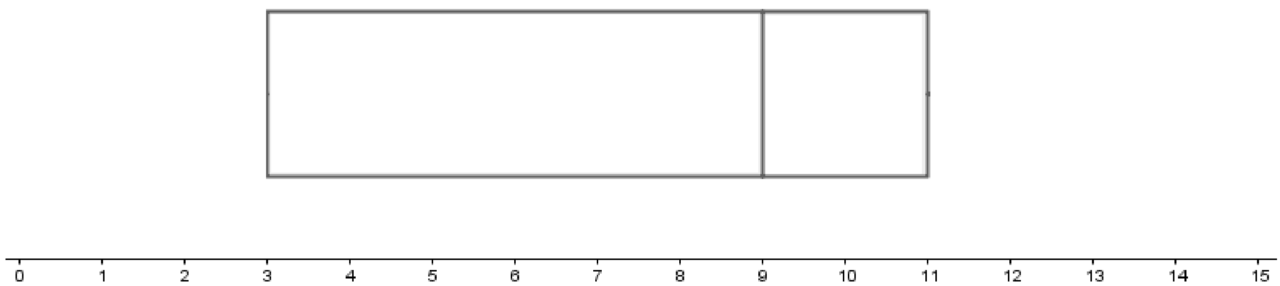
(i) Zuerst wählt man eine geeignete Skalierung der x-Achse. Das Diagramm sollte eine Querseite ausfüllen und alle gegebenen Werte abdecken.



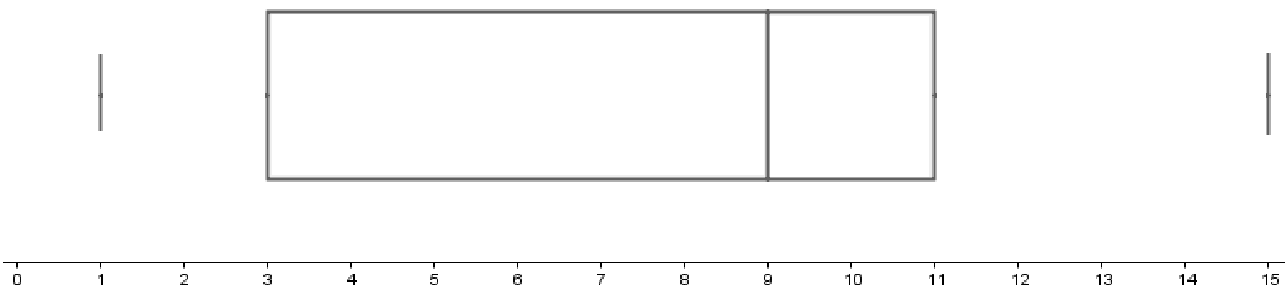
(ii) Danach werden die Quartile mit einem längeren Strich (2 cm) oberhalb der x-Achse markiert.



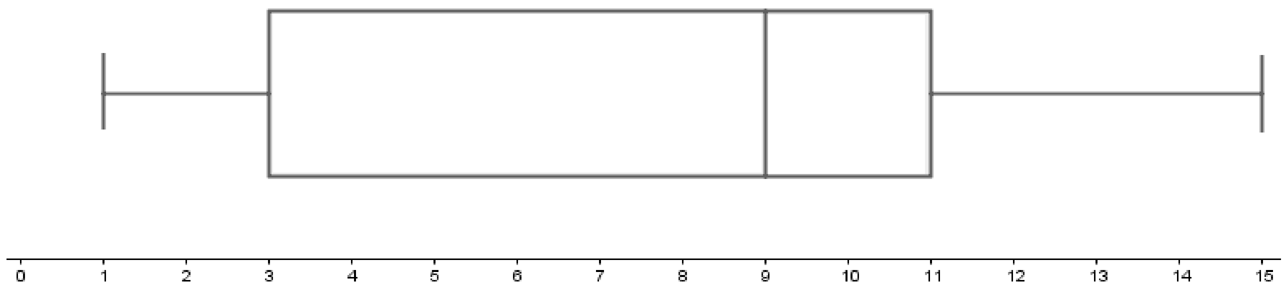
(iii) Die Quartilsmarkierungen werden zu einer Box verbunden.



(iv) Minimum und Maximum werden mit einem kleinen Strich (1 cm) markiert.



(v) Die kurzen Striche für das Minimum und das Maximum werden parallel zur x-Achse mit der Box verbunden.



Das entstandene Diagramm nennt man Kastenschaubild oder Boxplot.
Es liefert eine Übersicht über die Verteilung der Daten. Im vorliegenden Fall kann man sagen, dass die Daten im Bereich [9 ; 11] sehr dicht beieinander liegen, im Bereich [3 ; 9] jedoch weniger.
Allgemein gilt:
(Mindestens) 25% der Werte liegen in den Intervallen:

[min ; Q_1] [Q_1 ; m] [m ; Q_3] [Q_3 ; max]

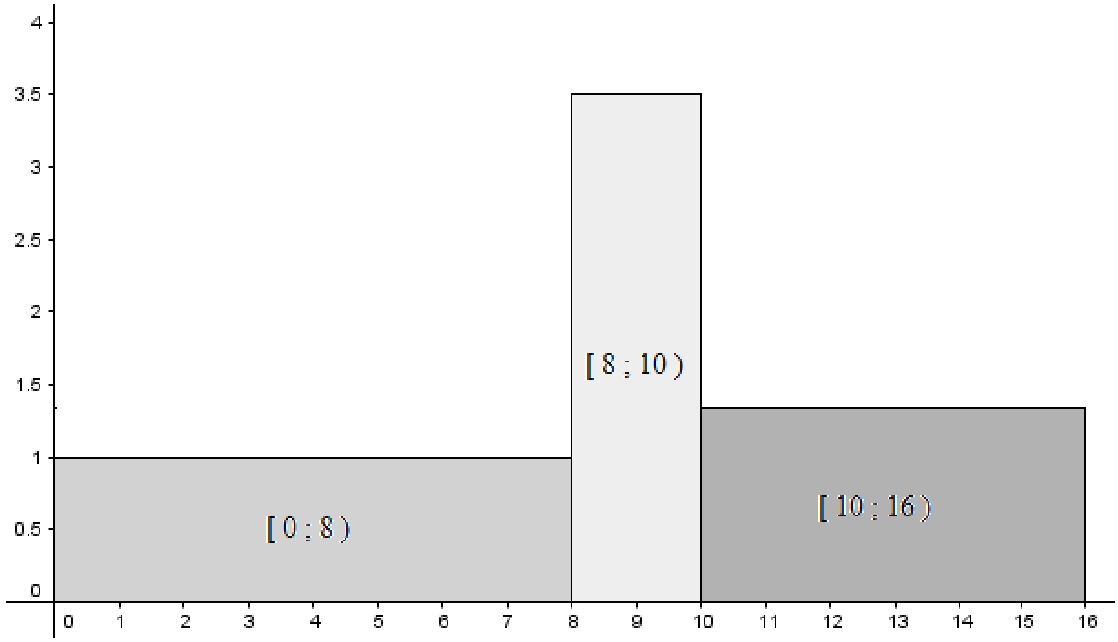
27) Histogramm
Diese Art der Darstellung ähnelt dem Säulendiagramm. Es berücksichtigt allerdings die Klassenbreite und gibt ähnlich wie das Kastenschaubild eine Übersicht über die Verteilung der Daten. Mit wenig Aufwand kann auch die absolute Häufigkeit errechnet werden.
Wir wählen hierfür eine neue Klasseneinteilung, in der die Klassen nicht gleich groß sind.

	Absolute Häufigkeit	Klassenbreite	Rechteckshöhe
1. Klasse: [0 ; 8)	8	8	8/8 = 1
2. Klasse: [8 ; 10)	7	2	7/2 = 3,5
3. Klasse: [10 ; 16)	8	6	8/6 ≈ 1,3

Die Klassenbreite ergibt sich aus der Anzahl ganzzahliger Werte einer Klasse.
Zum Beispiel: $[\,0\,;\,8\,)_Z := \{0, 1, 2, 3, 4, 5, 6, 7\} \rightarrow$ Anzahl: 8
$[\,8\,;\,10\,)_Z := \{8, 9\} \rightarrow$ Anzahl: 2
$[\,10\,;\,16\,)_Z := \{10, 11, 12, 13, 14, 15\} \rightarrow$ Anzahl: 6

Die Rechteckhöhe wird als Quotient der absoluten Häufigkeit und der Klassenbreite berechnet.
Zum Beispiel: [0 ; 8): 8/8 = 1
[8 ; 10): 7/2 = 3,5
[10 ; 16): 8/6 ≈ 1,3

Bei der Zeichnung sollte die x-Achse an die Werte der Urliste angepasst sein, die y-Achse an die Rechteckhöhe. Die Rechteckbreite (x-Achse) entspricht der Klassenbreite, die Rechteckhöhe spricht für sich selbst.



Mit einem Histogramm kann auf einem Blick festgestellt werden, wo die Daten am dichtesten zusammenliegen, hier nämlich im Intervall [8 ; 10).
Die absolute Häufigkeit kann ganz einfach errechnet werden, indem man die Klassenbreite und die Rechteckhöhe multipliziert. Klassenbreite mal Rechteckhöhe:
Zum Beispiel: [0 ; 8): $8 \cdot 1 = 8$; [8 ; 10): $2 \cdot 3{,}5 = 7$; [10 ; 15]: $6 \cdot 1{,}\dot{3} = 8$
→ hier sieht man auch warum das Intervall bis exklusive 16 gezogen wurde. Es würde sich sonst nicht mit der Rechteckbreite ausgehen. Man muss bedenken, das im Intervall [0 ; 15] 16 ganze Zahlen liegen.

Übung: Gegeben sind zwei Datensätze. Arbeite die Punkte 16) bis 27) der Anleitung für die Datensätze durch.
a) Das Körpergewicht in kg der Spieler einer Jugend-Fußballmannschaft wird erhoben. Die Daten wurden in folgender Tabelle dargestellt.

63	65	85	83	74	72	67	78	80	65	69	88	64	63	76	79

→ Klassen: [60 ; 65), [65 ; 75), [75 ; 90)

b) Bei einer Überprüfung der Wurfleistung der Spieler einer Basketballmannschaft wirft jeder 30 Freiwürde. Die Anzahl der Treffer wird notiert und in einer Tabelle dargestellt.

17	13	5	8	6	19	12	24	12	11	13	8	10	12	15	22	28	9	22

→ Klassen: [0 ; 10), [10 ; 15), [15 ; 20), [20 ; 30)