

Histogramme

Arbeitsblatt

Die Häufigkeitsverteilung von Merkmalen, wie Masse, Länge oder Zeit, kann durch Histogramme grafisch dargestellt werden. Eigenschaften und Konstruktion des Histogramms sowie der Zusammenhang mit der Wahrscheinlichkeitsdichte einer stetigen Zufallsvariablen werden in diesem Arbeitsblatt erklärt.



Eigenschaften eines Histogramms

- Im Histogramm wird die Häufigkeitsverteilung metrisch skalierten Merkmale, wie Masse, Länge, Anzahl, Zeit, Geschwindigkeit, Temperatur etc., dargestellt.
- Der Wertebereich der erfassten Daten wird in Klassen eingeteilt. Dabei können die Klassen gleich breit sein oder unterschiedliche Breite aufweisen. Die Intervalle der Klassen grenzen aneinander. Die beiden Klassen am linken und rechten Rand müssen geschlossen sein.
- Die relative Häufigkeit jeder Klasse wird mithilfe eines Rechtecks dargestellt.
- Der Flächeninhalt des Rechtecks ist proportional zur relativen Häufigkeit der jeweiligen Klasse.

Konstruktion eines Histogramms

Aus den Eigenschaften ergibt sich daher folgende Vorgangsweise.

- Die Anzahl der Klassen richtet sich nach der Anzahl n der Messungen bzw. der erhobenen Daten. Als Faustregel kann \sqrt{n} dienen, wobei die Anzahl der Klassen nicht kleiner als 5 und nicht größer als etwa 20 sein sollte.
- Die Intervallgrenzen der gewählten Klassen werden auf der x-Achse (Abszisse) auf einer Skala eingetragen.
- Bei Klassen unterschiedlicher Breite wird keine y-Achse (Ordinate) angegeben, da die Höhe des Rechtecks nicht der relativen Häufigkeit entspricht. Die relative Häufigkeit wird beim jeweiligen Rechteck angegeben.
- Bei Klassen gleicher Breite kann eine y-Achse zum Abtragen der relativen Häufigkeiten verwendet werden, da die Höhe der Rechtecke der jeweiligen relativen Häufigkeit entspricht.
- Die Rechtecke grenzen entsprechend den Intervallen (Klassen) ohne Abstand aneinander. Im Gegensatz dazu werden Säulendiagramme meist mit Abstand gezeichnet.
- Die Höhe eines Rechtecks ergibt sich aus dem Quotienten von relativer Häufigkeit und Klassenbreite.

Beispiel:

Von der Statistik Austria wurde erhoben, wie viele Kinder bzw. junge Menschen in Österreich sich im Jahr 2010 bzw. 2011 in welcher Bildungseinrichtung befanden. Die Auswertung der Daten ergab:

Bildungseinrichtung	ungefähres Alter	Anzahl (gerundet)
Kinderkrippe	0 bis 2 Jahre	31 000
Kindergarten	3 bis 5 Jahre	200 000
Volksschule (1. bis 4. Schulstufe)	6 bis 9 Jahre	322 000
Sekundarstufe I (5. bis 8. Schulstufe)	10 bis 13 Jahre	362 000
Sekundarstufe II (9. bis 13. Schulstufe)	14 bis 20 Jahre	440 000

Die Altersangaben sind auf folgende Weise zu lesen: 3 bis 5 Jahre bedeutet vom Beginn des 3. Lebensjahres bis zum vollendeten 5. Lebensjahr.

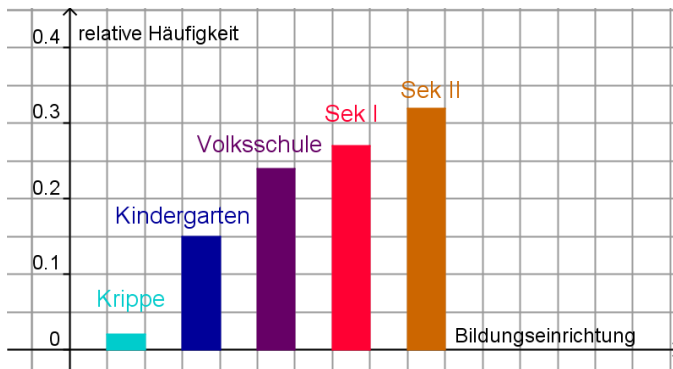
Erstelle zu den relativen Häufigkeiten ein Säulendiagramm, in dem nur die Bezeichnungen der Bildungseinrichtungen verwendet werden, und erstelle ein Histogramm, das die Altersklassen verwendet.

Quelle:
 STATISTIK AUSTRIA: Kindertagesheimstatistik 2010/11. Erstellt am 22.06.2011.
 STATISTIK AUSTRIA: Schulstatistik 2009/10. Erstellt am 29.11.2010.

Zunächst werden die relativen Häufigkeiten berechnet.

Bildungseinrichtung	Anzahl	rel. Häufigkeit
Kinderkrippe	31 000	0,02
Kindergarten	200 000	0,15
Volksschule (1. bis 4. Schulstufe)	322 000	0,24
Sekundarstufe I (5. bis 8. Schulstufe)	362 000	0,27
Sekundarstufe II (9. bis 13. Schulstufe)	440 000	0,32

Werden nur die Bezeichnungen der Bildungseinrichtungen verwendet, also ein qualitatives Merkmal, dann werden die Daten in einem Säulendiagramm dargestellt. Die Höhe der Säule entspricht der relativen Häufigkeit. Die Säulen können mit Abstand gezeichnet werden.



Wird ein quantitatives Merkmal verwendet, das metrisch skalierbar ist, wie in diesem Beispiel das Alter, dann werden die Daten zu Klassen zusammengefasst und können in einem Histogramm dargestellt werden. Durch den Untersuchungsgegenstand sind 5 unterschiedlich breite Altersklassen vorgegeben.

Der Flächeninhalt des jeweiligen Rechtecks muss proportional zur relativen Häufigkeit sein.

$$\text{Flächeninhalt} = \text{Breite} \cdot \text{Höhe}$$

$$\text{relative Häufigkeit} \sim \text{Klassenbreite} \cdot \text{Höhe des Rechtecks}$$

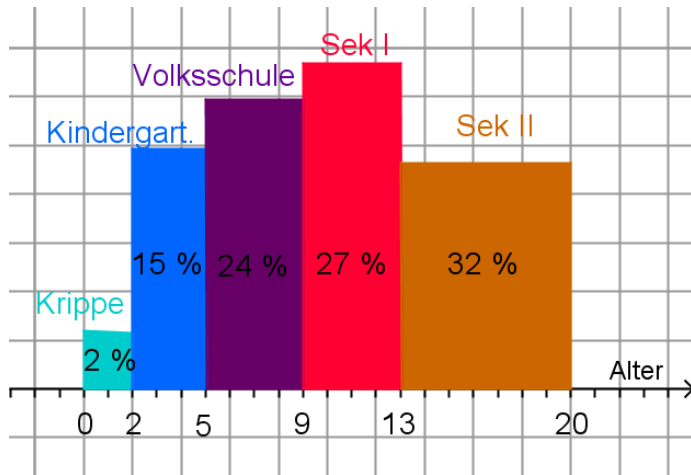
$$\text{Höhe des Rechtecks} \sim \frac{\text{relative Häufigkeit}}{\text{Klassenbreite}}$$

Mithilfe von relativer Häufigkeit und Klassenbreite wird die jeweilige Rechteckshöhe ermittelt.

Die Summe der Rechtecksflächen ergibt 1.

ungefähres Alter	Klassenbreite	rel. Häufigkeit	Rechteckshöhe
0 bis 2 Jahre	2	0,02	0,011
3 bis 5 Jahre	3	0,15	0,049
6 bis 9 Jahre	4	0,24	0,059
10 bis 13 Jahre	4	0,27	0,067
14 bis 20 Jahre	7	0,32	0,046

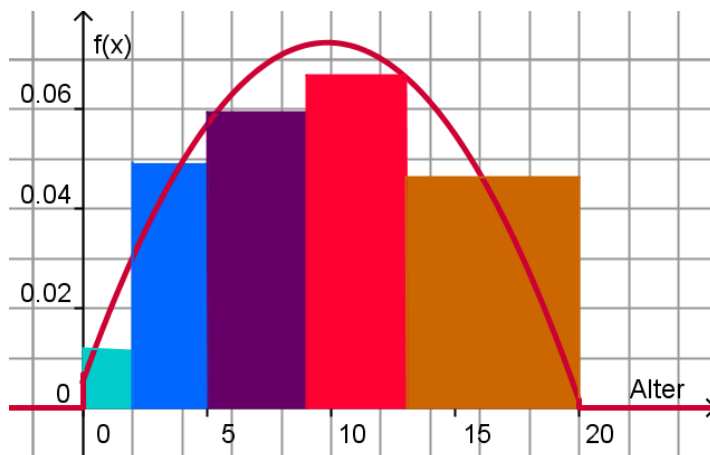
Im Histogramm werden nun entsprechende Rechtecke ohne Abstand gezeichnet und die x-Achse auf geeignete Weise skaliert. Die y-Achse wird nicht dargestellt.



Zusammenhang mit der Wahrscheinlichkeitsdichte einer stetigen Zufallsvariablen

Das Histogramm steht in engem Zusammenhang mit der Wahrscheinlichkeitsdichte einer stetigen Zufallsvariablen. Vergleiche:

Histogramm	Wahrscheinlichkeitsdichtefunktion $f(x)$
Der Wertebereich der Messwerte ist metrisch skalierbar und umfasst ein geschlossenes Intervall reeller Zahlen.	Die Zufallsvariable ist stetig.
Die Summe aller Rechtecksflächen ergibt 1.	Der Flächeninhalt zwischen x-Achse und $f(x)$ beträgt 1.
Die Höhen der Rechtecksflächen sind größer als 0 oder gleich 0.	$f(x) \geq 0$



Als Wahrscheinlichkeitsdichtefunktion wurde hier die Funktion $f(x) = -0,0007x^2 + 0,0138x + 0,0054$ im Intervall $[0; 20]$ verwendet.

Aufgaben

- 1 In einem Land wird an verschiedenen Wetterstationen morgens um 7 Uhr die Lufttemperatur in Grad Celsius gemessen. Es ergeben sich folgende Daten:

4,3°	-1,5°	0,7°	5,0°	-4,4°	0,0°	-12,1°	3,0°	-5,0°	-2,0°
7,1°	-4,1°	4,3°	6,6°	-3,9°	-1,9°	-10,9°	-7,0°	6,3°	-5,6°

Ermittle eine geeignete Klasseneinteilung: Wähle dabei ganzzahlige Intervallgrenzen. Die beiden Klassen am linken und rechten Rand sollen gerade die beiden äußersten Werte erfassen und werden daher eventuell schmaler als die anderen Klassen sein. Ermittle die absolute und relative Häufigkeit für die jeweilige Klasse. Erstelle ein Histogramm. Begründe, warum deine Darstellung die Eigenschaften eines Histogramms erfüllt.

- 2 Die folgende Tabelle gibt an, wie viele Mobiltelefone pro 100 Einwohner in europäischen Ländern verwendet werden. Ermittle eine geeignete Klasseneinteilung und die Anzahl der Länder in der jeweiligen Klasse. Bestimme mithilfe der absoluten Häufigkeiten die entsprechenden relativen Häufigkeiten und stelle die Daten in einem Histogramm dar.

Rang	Land	Anzahl
1	Estland	209
2	Italien	156
3	Bulgarien	147
4	Finnland	147
5	Luxemburg	146
6	Österreich	143
7	Portugal	142
8	Tschechien	140
9	Litauen	140
10	Dänemark	135
11	Kroatien	134
12	Serbien	134
13	Vereinigtes Königreich	132
14	Deutschland	128
15	Niederlande	127
16	Schweden	126
17	Spanien	126
18	Griechenland	124
19	Schweiz	122
20	Ukraine	121
21	Belgien	119

Rang	Land	Anzahl
22	Ungarn	119
23	Irland	116
24	Polen	116
25	Norwegen	114
26	Albanien	114
27	Rumänien	114
28	Island	114
29	Montenegro	112
30	Slowenien	105
31	Malta	104
32	Liechtenstein	101
33	Slowakei	101
34	Lettland	101
35	Frankreich	95
36	Mazedonien	94
37	Türkei	82
38	San Marino	79
39	Andorra	77
40	Bosnien und Herzegowina	71
41	Monaco	70

Quelle: www.indexmundi.com unter Country Comparison > Telephones - mobile cellular per capita (15.11.2011). Die Daten basieren auf CIA World Factbook vom 1.1.2011, <https://www.cia.gov/library/publications/the-world-factbook/>.

- 3 Suche in Zeitschriften, Büchern oder auf Internetseiten Beispiele für korrekte Histogramme und für Darstellungen, die fälschlicherweise als Histogramm bezeichnet werden.