

Statistik

Dr. Thomas Zehrt

Merkmale und Häufigkeitsverteilung

Motivation

In der heutigen Zeit fällt jeden Tag eine unvorstellbare Menge von Daten an, z.B. Kurse von Aktien, Obligationen und Staatsanleihen, Arbeitslosenzahlen, Arbeitslosenquoten, Preisindizes usw. Eine der Hauptaufgaben der **beschreibenden Statistik** ist es nun, solche grossen unübersichtlichen Datenmengen (durch Computereinsatz) aufzuarbeiten und durch wenige aussagekräftige Kenngrössen möglichst gut zu beschreiben (SMI, DAX, Dow-Jones-Index). Natürlich erhoffen wir, dass uns diese (aufbereiteten) Daten zumindest hilfreich sind, um gewisse Fragen zu beantworten:

- Welchen Zustand haben wir heute erreicht? (z.B. Wie hoch ist das Bruttosozialprodukt?)
- Welchen Zustand hatten wir in der Vergangenheit und können wir den zukünftigen Zustand prognostizieren? (z.B. Ist das Bruttosozialprodukt in den letzten Jahren gewachsen und wie könnte es sich in den nächsten Jahren entwickeln?)
- Welche Auswirkungen hatten oder haben bestimmte Massnahmen auf diesen Zustand? (z.B. Wie hat oder wird eine Steuersenkung das Bruttosozialprodukt beeinflussen bzw. beeinflussen?)

Benötigtes Schulwissen

- Darstellung von Mengen in der Form $\{ x \in \mathbb{R} \mid \text{Eigenschaft, die } x \text{ haben muss} \}$
- Bruchrechnung und Prozentrechnung
- Summenzeichen $\sum_{j=1}^m f_j = f_1 + f_2 + \dots + f_m$, also z.B. falls $m = 4$ und $f_1 = 2$, $f_2 = 5$, $f_3 = -2$, $f_4 = 0.5$ so ist $\sum_{j=1}^m f_j = 2 + 5 + (-2) + 0.5 = 5.5$.
- halboffene (d.h. einer der Randpunkte gehört zur Menge, der andere nicht), abgeschlossene und offene Intervalle als Teilmengen der reellen Zahlen, also z.B.

$$[1, 3) = \{ x \in \mathbb{R} : 1 \leq x < 3 \}$$

$$[1, 3] = \{ x \in \mathbb{R} : 1 \leq x \leq 3 \}$$

$$(1, 3) = \{ x \in \mathbb{R} : 1 < x < 3 \}$$

- Treppenfunktionen

1 Grundbegriffe

Die statistische Masse besteht aus allen Elementen (Personen: Kunden, Mitarbeiter, Studierende, ...; Sachen: Maschinen, Computer, ...; Institutionen: Betriebe, Länder, ...; Ereignisse: Geburten, Todesfälle) denen das Interesse des Statistikers gilt. Im mathematischen Sinne ist das eine Menge und die einzelnen Elemente heissen statistische Elemente oder Merkmalssträger.

Wichtig: Vor jeder statistischen Tätigkeit muss die statistische Masse in räumlicher, zeitlicher und sachlicher Hinsicht exakt definiert werden.

Wie der Name schon sagt, interessieren uns an den Merkmalssträgern gewisse Eigenschaften oder Merkmale und die Werte oder Zustände, die ein Merkmal annehmen kann, heissen Merkmalsausprägungen. Ein sinnvolles Merkmal muss mindestens zwei verschiedene Ausprägungen haben.

Beispiel 1.1 *Eine Computerzeitschrift testet verschiedene Drucker, d.h. die statistische Masse sind die getesteten Drucker. Es interessieren uns nun die folgenden Merkmale:*

- *Hersteller (Namen)*
- *Bezeichnung des Gerätes (alle Bezeichnungen)*
- *Preis (0-10000 CHF)*
- *Gewicht (0-10 kg)*
- *Seitenzahl pro Minute (0-100 Seiten)*
- *Gesamturteil (sehr gut, gut, mittel, schlecht, sehr schlecht)*

In Klammern stehen die möglichen Merkmalsausprägungen.

Merkmalskategorien

- **qualitatives Merkmal**

Das sind Merkmale, die keine physikalische (Mass)einheit benötigen. Man unterscheidet hier weiter zwischen:

- **qualitativ-nominalen Merkmalen**

Hier sind Merkmalsausprägungen nur dem Namen nach zu unterscheiden, drücken aber keinerlei Wertung oder Intensität aus.

- **qualitativ-ordinalen Merkmalen**

Hier unterscheiden sich die Merkmalsausprägungen nicht nur hinsichtlich des Namens, sondern können zusätzlich noch in eine (sinnvolle) Rangordnung gebracht werden.

- **quantitatives Merkmal**

Hier braucht die Nennung des Merkmals eine (physikalische) Einheit, um sinnvoll zu sein. Man unterscheidet hier weiter zwischen:

- **quantitativ-diskreten Merkmalen**

Diese Merkmale haben einen Wert, der durch einen Abzählprozess entsteht und zwischen den einzelnen Ausprägungen können keine Werte angenommen werden.

- **quantitativ-stetigen Merkmalen**

Diese Merkmale werden durch Messungen gewonnen und können (zumindest theoretisch) jeden Wert innerhalb eines Messbereiches annehmen.

Beispiel 1.2 Für unsere Drucker (vgl. Beispiel 1.1) sind „Hersteller“, „Bezeichnung“ und „Geamturteil“ qualitative Merkmale, denn sie tragen keine physikalische Einheit. „Geamturteil“ ist auch noch ordinal, denn die Werte können in eine sinnvolle Reihenfolge gebracht werden. Alle anderen Merkmale sind quantitativ. „Preis“ und „Gewicht“ sind stetig und „Seitenzahl pro Minute“ ist diskret. (Streng genommen ist auch „Preis“ diskret!)

Skalierung von Merkmalen Eine weitere Möglichkeit zwischen Merkmalen zu unterscheiden ist die Untersuchung der Skala, auf der sie gemessen werden. Wir unterscheiden zunächst die folgenden drei Skalenniveaus:

- **Nominalskala**

Nominalskalierte Merkmalsausprägungen können nicht sinnvoll geordnet werden (Merkmal ‚Geschlecht‘).

- **Ordinal- oder Rangskala**

Die Merkmalsausprägungen können gemäss ihrer Intensität geordnet werden. Eine Interpretation der Rangordnung ist möglich, Abstände zwischen Ausprägungen können aber nicht interpretiert werden (Merkmal ‚Schulnote‘).

- **Metrische Skala**

Unter den Merkmalsausprägungen kann eine Rangordnung definiert werden **und** Abstände zwischen Ausprägungen können gemessen und interpretiert werden.

Weitere Unterteilung:

- **Intervallskala**

Nur Differenzenbildung zweier Ausprägungen sind zulässig; nur Abstände können verglichen werden (Merkmal ‚Temperatur‘).

- **Verhältnisskala**

Es existiert zusätzlich ein natürlicher Nullpunkt. Quotientenbildung von Merkmalsausprägungen ist zulässig und sinnvoll interpretierbar (Merkmal ‚Geschwindigkeit‘).

- **Absolutskala**

Es kommt zusätzlich eine **natürliche** (Mass)einheit hinzu (Merkmal ‚Semesteranzahl‘).

Stichproben Werden alle Elemente einer statistischen Masse in die Erhebung einbezogen, liegt eine Total- oder Vollerhebung vor, ansonsten eine Teilerhebung oder Stichprobe. Elemente in Stichproben können zufällig ausgewählt werden (zufällige Stichprobe) oder bewusst zusammengesetzt werden, so dass in der Stichprobe die Werte gewisser Merkmale mit der gleichen relativen Häufigkeit repräsentiert sind wie in der gesamten statistischen Masse (repräsentative Stichprobe). Natürlich benötigt man Informationen über die gesamte statistische Masse, um eine repräsentative Stichprobe auszuwählen.

Informationen über die gesamte statistische Masse	\implies	repräsentative Stichprobe
zufällige Stichprobe	$? \implies ?$	Informationen über die gesamte statistische Masse

2 Eindimensionale Häufigkeitsverteilung

Das Wort eindimensional bezieht sich hier auf die Untersuchung eines einzelnen Merkmals.

Direkt bei der Datenerhebung entsteht zunächst die sogenannte Urliste. Sie ist im allgemeinen sehr unübersichtlich und nicht dazu geeignet, sich einen Überblick über die Merkmalsausprägung zu verschaffen. Deshalb stellt man zunächst die Häufigkeitsverteilung des Merkmals X in Form einer unklassierten oder klassierten Häufigkeitstabelle dar.

Definition 2.1 Sei n die Anzahl der statistischen Elemente, welche von 1 bis n durchnummeriert seien. Dann ist für alle $i = 1, \dots, n$:

$$x_i := \text{Ausprägung des Merkmals } X \text{ bei Element } i.$$

Ist nun $m (\leq n)$ die Anzahl der verschiedenen Merkmalsausprägungen von X in der statistischen Masse, dann sei für alle $j = 1, \dots, m$:

$$a_j := j\text{-te Ausprägung des Merkmals } X.$$

Die absolute Häufigkeit h_j und die relative Häufigkeit f_j sind erklärt durch:

$$h_j := \text{Anzahl der } x_i \text{ mit der Ausprägung } a_j,$$

$$f_j := \frac{h_j}{n}.$$

Satz 1 Es gelten die folgenden Beziehungen:

$$0 \leq h_j \leq n \quad \text{und} \quad \sum_{j=1}^m h_j = n$$

$$0 \leq f_j \leq 1 \quad \text{und} \quad \sum_{j=1}^m f_j = 1.$$

Bei einem (quantitativ-)diskreten Merkmal können wir meist alle erhaltenen Merkmalsausprägungen direkt auflisten. In diesem Fall sprechen wir von einer unklassierten Häufigkeitsverteilung. Ihre graphische Darstellung liefert das Stabdiagramm. Die Merkmalsausprägungen werden dabei immer an der waagerechten Axe angetragen, die Länge der Stäbe entspricht der Häufigkeit. Tatsächlich ist es hier (fast) egal, ob man die absolute oder relative Häufigkeit betrachtet. Der Unterschied drückt sich nur in der Beschriftung der senkrechten Axe aus.

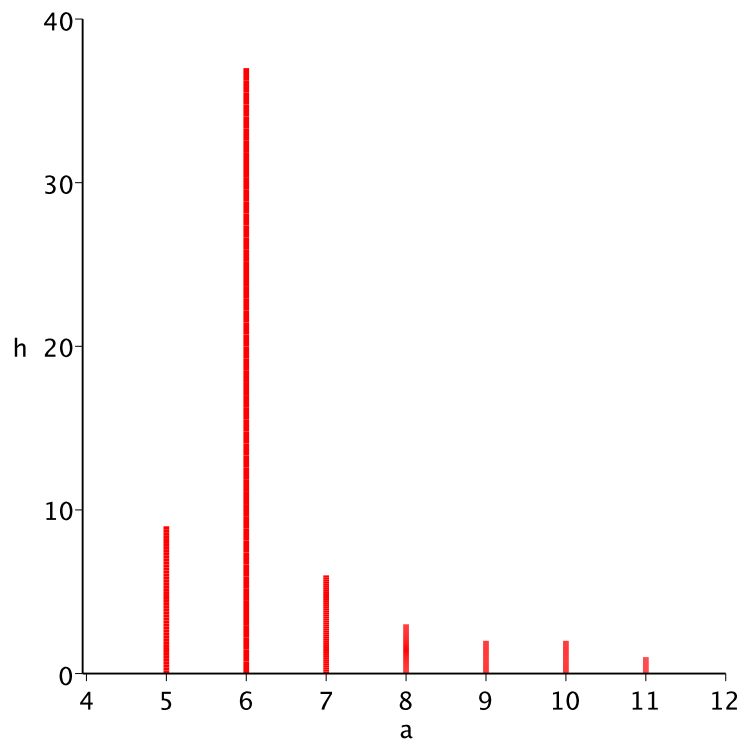
Beispiel 2.1 Bei einer Befragung von 60 erfolgreichen Studienabsolventen wird das Merkmal X : „Anzahl der Semester bis zum Bachelor“ untersucht. Es entsteht zunächst die folgende Urliste.

$x_1 = 6 \ 6 \ 7 \ 5 \ 6 \ 6 \ 6 \ 6 \ 6 \ 5 \ 6 \ 8 \ 6 \ 9 \ 6 \ 6 \ 7 \ 6 \ 5 \ 5$
 $6 \ 6 \ 8 \ 5 \ 6 \ 6 \ 10 \ 6 \ 6 \ 5 \ 6 \ 6 \ 6 \ 6 \ 6 \ 7 \ 6 \ 5 \ 10$
 $7 \ 6 \ 7 \ 6 \ 6 \ 6 \ 6 \ 6 \ 6 \ 5 \ 6 \ 8 \ 6 \ 9 \ 6 \ 6 \ 7 \ 6 \ 5 \ 11 = x_{60}$

Natürlich ist $n = 60$ und das Merkmal X hat die sieben Ausprägungen $a_1 = 5$, $a_2 = 6$, $a_3 = 7$, $a_4 = 8$, $a_5 = 9$, $a_6 = 10$ und $a_7 = 11$. Wir erstellen die Häufigkeitstabelle:

Semesteranzahl a_j	Strichliste	Häufigkeit	
		absolut h_j	relativ f_j
5		9	0.1500
6		37	0.6167
7		6	0.1000
8		3	0.0500
9		2	0.0333
10		2	0.0333
11		1	0.0167
Σ		60	1

In einem Stabdiagramm können wir nun die (absoluten bzw. relativen) Häufigkeiten graphisch darstellen.



Beispiel 2.2 *An einer Klausur in Stochastik an der Universität nahmen 365 Studenten und Studentinnen teil. Es waren insgesamt 24 Punkte zu erreichen. Die statistische Masse sind also diese 365 Studenten und das Merkmal ist die erreichte Punktzahl. Nach der Korrektur gab es folgende Punkteverteilung:*

<i>Punkte a_j</i>	<i>absolute Häufigkeit h_j</i>	<i>relative Häufigkeit f_j</i>
0	0	0
1	6	0.016438356
2	2	0.005479452
3	7	
4	5	
5	6	
6	7	
7	9	
8	6	
9	10	
10	15	
11	25	
12	12	
13	25	
14	0	
15	55	
16	35	
17	21	
18	30	
19	27	
20	18	
21	20	
22	10	
23	12	
24	2	

Es fällt die seltsame Lücke bei 14 Punkten auf! Tatsächlich war die notwendige Punktzahl für das Bestehen der Klausur 15 Punkte und offensichtlich wurden Studenten und Studentinnen knapp unterhalb der Mindestpunktzahl “gnadenhalber” hochgestuft.

Beispiel 2.3 Bei einem **Buchstaben-Cäsar** handelt es sich um ein Verschlüsselungsverfahren, bei dem jeder Buchstabe eines Klaralphabets (z.B. das deutsche Alphabet) durch einen Buchstaben im Geheimalphabet (z.B. das deutsche Alphabet) ersetzt wird. Ein wichtiges Hilfsmittel zur Entzifferung von Buchstaben-Cäsaren sind Statistiken der relativen Häufigkeit der Buchstaben des deutschen Alphabets. Die nachfolgenden Ergebnisse (die relative Häufigkeit der Buchstaben) sind aus 40 000 Buchstaben von deutschen Texten verschiedenen Inhalts ermittelt worden.

<i>e</i>	0.1798	<i>m</i>	0.0247
<i>n</i>	0.1106	<i>w</i>	0.0203
<i>i</i>	0.0797	<i>o</i>	0.0200
<i>s</i>	0.0748	<i>b</i>	0.0177
<i>r</i>	0.0642	<i>f</i>	0.0123
<i>a</i>	0.0596	<i>z</i>	0.0121
<i>t</i>	0.0555	<i>k</i>	0.0112
<i>d</i>	0.0522	<i>v</i>	0.0076
<i>u</i>	0.0487	<i>p</i>	0.0059
<i>h</i>	0.0461	<i>j</i>	0.0006
<i>g</i>	0.0325	<i>q</i>	0.0001
<i>l</i>	0.0319	<i>x</i>	0.0001
<i>c</i>	0.0317	<i>y</i>	0.0001

Wer Spass daran hat kann versuchen, die folgende Aufgabe zu lösen. Die obige Tabelle ist dabei zwar hilfreich, aber ohne Probiere und Knobeln werden Sie wohl nicht ans Ziel kommen.

Aufgabe 2.1 Versuchen Sie mit Hilfe der Tabelle in Beispiel 2.3 den folgenden Geheimtext zu entschlüsseln (dabei sind Umlaute wie z.B. ü als ue geschrieben):

„abcc dbe afcd dbg ahcdbig abjk lhmbc dfb bfcbc nhmbec mcd dfb hcdbebc afcdnmbjibe“.

Lösung:

Stetige Merkmale Bei einem stetigen Merkmal erhalten wir in der Regel sehr viele verschiedene Merkmalsausprägungen und bei genügend grosser Messgenauigkeit ist sogar zu erwarten, dass alle n Elemente verschiedene Werte x_i haben. Für sehr grosse n ist es deshalb meist sinnvoll, die erhaltenen Merkmalsausprägungen zu Klassen zusammenzufassen (d.h. in der Gestalt „von ... bis ...“). In diesem Fall sprechen wir von einer klassierten Häufigkeitsverteilung.

Empfehlungen:

1. Vermeide nach unten und nach oben offene Klassen am Anfang und am Ende.
2. Die Anzahl der Klassen sollte den Wert \sqrt{n} nicht überschreiten.

Natürlich zahlt man einen Preis für den Gewinn an Übersichtlichkeit, die man durch die Klasseneinteilung gewinnt: Informationsverlust, denn über die Verteilung der Werte innerhalb einer Klasse ist nichts mehr bekannt.

Definition 2.2 Gegeben sei ein stetiges Merkmal X mit Werten in einem Intervall $[a, b)$ und eine Folge von reellen Zahlen

$$a = a_1 < a_2 < \dots < a_m < a_{m+1} = b.$$

Das Intervall $[a_j, a_{j+1})$ heisst die j -te Klasse K_j .

Die absolute Häufigkeit h_j und die relative Häufigkeit f_j der Klasse K_j sind erklärt durch

$$h_j := \text{Anzahl der Ausprägungen } x \text{ mit } x \in K_j = [a_j, a_{j+1})$$

$$f_j := \frac{h_j}{n}.$$

Weiterhin ist

$$d_j := a_{j+1} - a_j \text{ die } \underline{\text{Klassenbreite}}$$

$$m_j := a_j + \frac{d_j}{2} = \frac{a_{j+1} + a_j}{2} \text{ die } \underline{\text{Klassenmitte}}$$

$$h_j^* := \frac{h_j}{d_j} \text{ die } \underline{\text{Klassendichte}}.$$

Die graphische Darstellung einer klassierten Häufigkeitsverteilung erfolgt mittels eines Histogramms. Dabei werden die Intervallgrenzen $a = a_1 < a_2 < \dots < a_m < a_{m+1} = b$ (oder eben die Klassen K_1, K_2, \dots, K_m) auf der horizontalen Achse abgetragen. Dann wird über jeder Klasse ein Rechteck errichtet, dessen Fläche proportional zur Häufigkeit der jeweiligen Klasse ist. Zwei wichtige Spezialfälle sind dabei:

- das absolute Histogramm: hier gilt für jede Klasse K_j die Bedingung:

$$\text{Rechtecksfläche über } K_j (= d_j \cdot \text{Höhe}) = h_j \quad \text{also} \quad \text{Höhe} = \frac{h_j}{d_j} = h_j^*$$

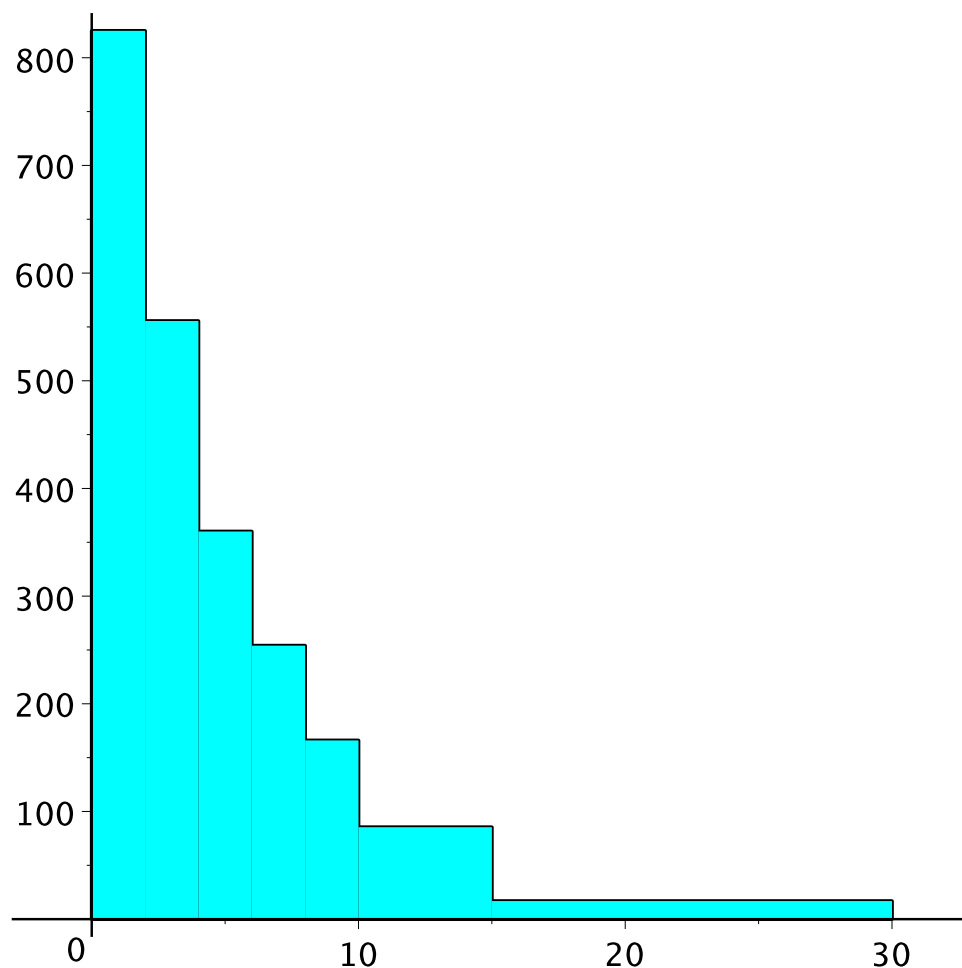
- das relative Histogramm: hier gilt für jede Klasse K_j die Bedingung:

$$\text{Rechtecksfläche über } K_j (= d_j \cdot \text{Höhe}) = f_j \quad \text{also} \quad \text{Höhe} = \frac{f_j}{d_j} = \frac{h_j^*}{n}$$

Beispiel 2.4 Von 5000 Telefongesprächen wurden an der zentralen Telefonanlage die Dauer in Minuten gemessen. Die folgende Tabelle gibt die klassierte absolute und relative Häufigkeitsverteilung an, wobei nicht alle Klassen gleich lang sind. Dabei bezeichnet $K_j = [a_j, a_{j+1})$ und die a_j sind in Minuten gegeben.

j	Klasse K_j	h_j	f_j	d_j	h_j^*	m_j
1	$[0, 2)$	1650	0.3300	2	825.0	1.0
2	$[2, 4)$	1111		2		
3	$[4, 6)$	720		2		
4	$[6, 8)$	508		2		
5	$[8, 10)$	332		2		
6	$[10, 15)$	427		5		
7	$[15, 30)$	252		15	16.8	22.5

Alle Gesprächsdauern (aller 5000 Gespräche) waren im Intervall $[0, 30)$ zu finden und wir bilden (willkürlich) die folgenden Klassen: $a_1 = 0 < 2 < 4 < 6 < 8 < 10 < 15 < 30 = a_7$.



3 Die empirische Verteilungsfunktion

In der Praxis interessiert man sich oft für Fragestellungen der folgenden Art:

- Welcher Anteil der statistischen Masse liegt, bezüglich eines fixierten Merkmals, unterhalb oder oberhalb einer bestimmten Grenze, oder zwischen zwei Grenzen?
- Welche Merkmalsausprägung x_p hat die Eigenschaft, dass der prozentuale Anteil p der statistischen Masse unterhalb x_p (und der Rest oberhalb dieses Parameters) liegt?

Diese Fragen lassen sich mit der empirischen Verteilungsfunktion effizient beantworten. Der Begriff empirisch bedeutet soviel wie: durch Beobachtung gewonnen.

Definition 3.1 *Die empirische Verteilungsfunktion F ist, für eine nicht in Klassen eingeteilte statistische Masse, gegeben durch*

$$F(x) := \sum_{a_j \leq x} f_j = \sum_{\substack{j \in \mathbb{N} \\ \text{s.d. } a_j \leq x}} f_j$$

Wir addieren also alle relativen Häufigkeiten f_j die zu Ausprägungen a_j gehören, die nicht rechts von x (links von x oder genau auf x) liegen.

Aufgabe 3.1 *Einhundert Geräte des gleichen Typs werden im Rahmen einer Qualitätskontrolle in einem bestimmten Zeitraum beobachtet und bei jedem Gerät die Anzahl der Ausfälle in einem Zeitraum gezählt (statistische Masse: die 100 Geräte, Merkmal: Anzahl der Ausfälle). Man erhält folgende Häufigkeitstabelle:*

Ausfälle	$a_1 = 2$	$a_2 = 3$	$a_3 = 4$	$a_4 = 5$	$a_5 = 6$	$a_6 = 7$	$a_7 = 8$	$a_8 = 9$
abs. Häufg.	$h_1 = 9$	$h_2 = 17$	$h_3 = 24$	$h_4 = 18$	$h_5 = 13$	$h_6 = 10$	$h_7 = 6$	$h_8 = 3$
rel. Häufg.	$f_1 = \frac{9}{100}$	$f_2 = \frac{17}{100}$	$f_3 = \frac{24}{100}$	$f_4 = \frac{18}{100}$	$f_5 = \frac{13}{100}$	$f_6 = \frac{10}{100}$	$f_7 = \frac{6}{100}$	$f_8 = \frac{3}{100}$

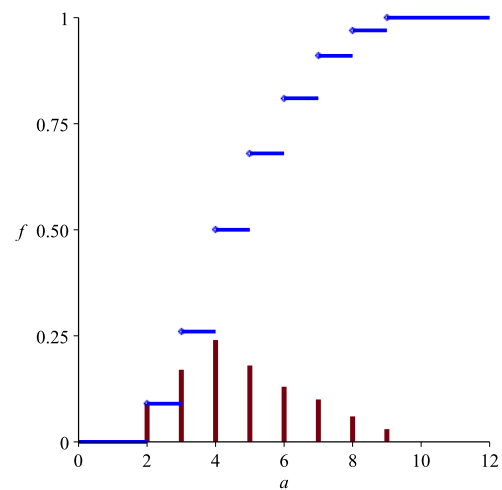
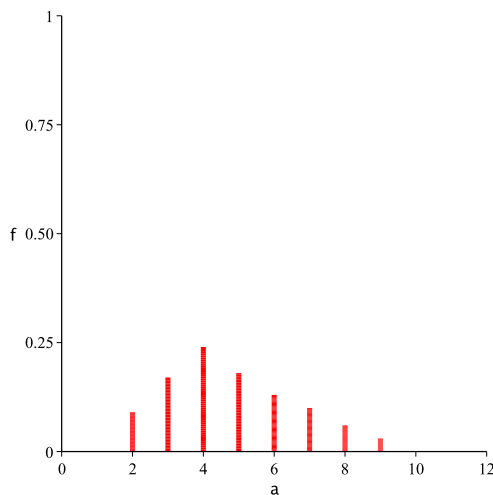
Bestimmen Sie die empirische Verteilungsfunktion.

Lösung:

1. Für jede reelle Zahl x ist (relativ) zu zählen, wieviele Ausprägungen man links von x und auf x findet.

$$F(x) := \begin{cases} 0 & = 0 & \text{für } x \in (-\infty, 2) \\ 9/100 & = 0.09 & \text{für } x \in [2, 3) \\ (9 + 17)/100 & = 0.26 & \text{für } x \in [3, 4) \\ (9 + 17 + 24)/100 & = 0.50 & \text{für } x \in [4, 5) \\ (9 + 17 + 24 + 18)/100 & = 0.68 & \text{für } x \in [5, 6) \\ (9 + 17 + 24 + 18 + 13)/100 & = 0.81 & \text{für } x \in [6, 7) \\ (9 + 17 + 24 + 18 + 13 + 10)/100 & = 0.91 & \text{für } x \in [7, 8) \\ (9 + 17 + 24 + 18 + 13 + 10 + 6)/100 & = 0.97 & \text{für } x \in [8, 9) \\ (9 + 17 + 24 + 18 + 13 + 10 + 6 + 3)/100 & = 1 & \text{für } x \in [9, \infty) \end{cases}$$

Stabdiagramm sowie Stabdiagramm und Graph von F (blau)



2. 18%
3. $F(4) = 0.5 = 50\%$
4. $1 - F(5) = 0.32 = 32\%$

4 Übungsaufgaben

1. Berechnen Sie

$$a) \sum_{k=-2}^2 (-1)^k \cdot k \cdot 2^k \quad b) \sum_{\substack{i \in \mathbb{N} \\ \text{s.d. } 3 < i < 7}} i^2 \quad c) \sum_{k=1}^{100} 4 \quad d) \sum_{k=0}^5 \left(\frac{3}{4}\right)^k$$

2. In der Praxis kommen oft Funktionen vor, die stückweise definiert sind, d.h. je nach dem wo man sich im Definitionsbereich befindet, werden andere Vorschriften zur Berechnung der Funktionswerte benutzt. Skizzieren Sie die Graphen der folgenden Funktionen per Hand und kontrollieren Sie dann Ihr Ergebnis mit Geogebra.

Machen Sie sich auch klar, welchen Wert die Funktionen an den Randpunkten der Teilintervalle annehmen.

$$f(x) := \begin{cases} -1 & \text{für } x < 0 \\ 0 & \text{für } x \in [0, 1) \\ 1 & \text{für } x \in [1, 2) \\ 4 & \text{für } x \geq 2 \end{cases} \quad g(x) := \begin{cases} 5 & \text{für } x < 0 \\ x^2 + 5 & \text{für } x \in [0, 2] \\ 0.5x + 8 & \text{für } x > 2 \end{cases}$$

$$h(x) := \begin{cases} -1 & \text{für } x \in [-1, 1] \\ |x| & \text{sonst} \end{cases} \quad i(x) := \begin{cases} 0 & \text{für } x < 0 \\ |x| & \text{für } x \in [0, 1) \\ |x + 1| & \text{für } x \in [1, 3) \\ 0 & \text{für } x \geq 3 \end{cases}$$

3. Sei $a_1 = 5$, $a_2 = 6$, $a_3 = 7$, $a_4 = 8$ sowie $f_1 = 1/8$, $f_2 = 3/8$, $f_3 = 2/8$, $f_4 = 2/8$ und

$$F(x) = \sum_{\substack{j \in \mathbb{N} \\ \text{s.d. } a_j \leq x}} f_j.$$

Berechnen Sie $F(1)$, $F(5)$, $F(5.5)$, $F(7)$, $F(7.8)$, $F(8)$ und $F(100)$.

4. Nennen Sie zu den folgenden Merkmalsträgern und Merkmalen jeweils einige sinnvolle Merkmalsausprägungen und geben Sie an, um welche Merkmalskategorie es sich handelt (qualitativ-nominal, qualitativ-ordinal, quantitativ-diskret oder quantitativ-stetig):

Merkmalsträger	Merkmal	einige Ausprägungen	Kategorie
Personen	Hochschulstudium (Ja/Nein)		
Personen	Familienstand		
Personen	Gewicht		
Hotels	Anzahl Sterne		
Personen	Alter in Jahren		
Personen	Alter (exakt)		
Betriebe	Anzahl Mitarbeiter		
Motoren	Leistung		

5. Die $n = 30$ Versuchstiere einer Population werden an einem Tag gewogen. Man erhält die folgende Urliste:

Lfd.Nr.	Gewicht in kg										
1 - 10	12.16	11.53	14.02	11.85	10.94	11.83	12.94	11.46	13.15	12.70	
11 - 20	10.88	13.24	14.04	10.95	14.78	12.39	13.69	11.82	14.28	12.96	
21 - 30	13.24	13.42	12.23	15.04	11.34	12.28	13.42	13.93	14.73	11.28	

- (a) Erstellen Sie, um einen besseren Überblick über die Verteilung der Merkmalsausprägungen zu haben, eine absolute klassierte Häufigkeitsverteilung mit den (gleich langen) Klassen: 10 bis unter 11 kg, 11 bis unter 12 kg usw.
- (b) Benennen Sie explizit alle Größen $m, a_1, \dots, a_{m+1}, K_1, \dots, K_m, h_1, \dots, h_m, f_1, \dots, f_m, d_1, \dots, d_m, m_1, \dots, m_m, h_1^*, \dots, h_m^*$.
- (c) Stellen Sie die absolute klassierte Häufigkeitsverteilung in einem Histogramm dar.
6. Der Besitzer eines Lebensmittelgeschäftes möchte wissen, ob es sich lohnt, täglich frische Milch in Flaschen zu beziehen oder nur H-Milch anzubieten. Dazu notiert er 100 Tage lang die Anzahl verkaufter Flaschen Frischmilch und erhielt folgende Urliste: 3, 4, 0, 0, 2, 3, 3, 1, 2, 2, 2, 5, 4, 5, 7, 7, 8, 8, 3, 3, 6, 7, 3, 2, 1, 7, 5, 5, 4, 1, 3, 8, 7, 5, 5, 4, 3, 3, 4, 4, 3, 3, 3, 2, 5, 5, 4, 3, 3, 3, 3, 4, 8, 7, 7, 5, 4, 1, 1, 3, 0, 0, 1, 0, 1, 0, 4, 3, 2, 1, 2, 1, 4, 0, 0, 1, 1, 1, 5, 0, 0, 2, 2, 3, 1, 3, 6, 6, 7, 6, 7, 4, 3, 2, 3, 1, 5, 3, 3, 2.
- (a) Geben Sie die Häufigkeitstabelle an.
- (b) Bestimmen Sie die empirische Verteilungsfunktion und stellen Sie diese graphisch dar.
- (c) Wie wird sich der Ladenbesitzer entscheiden, wenn er an mindestens 70% der Verkaufstage mindestens 2 Flaschen Frischmilch verkaufen **muss**?

Lösungen einiger Übungsaufgaben

1. a)

$$\begin{aligned} & \sum_{k=-2}^2 (-1)^k \cdot k \cdot 2^k \\ &= (-1)^{-2} \cdot (-2) \cdot 2^{-2} + (-1)^{-1} \cdot (-1) \cdot 2^{-1} + (-1)^0 \cdot 0 \cdot 2^0 + (-1)^1 \cdot 1 \cdot 2^1 + (-1)^2 \cdot 2 \cdot 2^2 \\ &= -0.5 + 0.5 + 0 - 2 + 8 = 6 \end{aligned}$$

b) $4^2 + 5^2 + 6^2$, c) 400 und d) 3.288

2. Geben Sie in der Geogebra-Eingabezeile folgende Befehle ein:

$$f(x) = \text{Wenn}(x < 0, -1, x < 1, 0, x < 2, 1, 4)$$

$$g(x) = \text{Wenn}(x < 0, 5, x \leq 2, x^2 + 5, 0.5 * x + 8)$$

$$h(x) = \text{Wenn}(-1 \leq x \leq 1, -1, \text{abs}(x))$$

$$i(x) = \dots$$

3. $F(1) = 0$, $F(5) = 1/8$, $F(5.5) = 1/8$, $F(7) = 6/8$, $F(7.8) = 6/8$, $F(8) = 1$ und $F(100) = 1$

4. -

5. (a)

Klasse von ... bis unter ... kg	abs. Häufigkeit
10-11	3
11-12	7
12-13	7
13-14	7
14-15	5
15-16	1

(b) $m = 6$, $a_1 = 10, \dots, a_7 = 16$, $K_1 = [10, 11), \dots, K_6 = [15, 16)$, $h_1 = 3, \dots, h_6 = 1$, $f_1 = 3/30, \dots, f_6 = 1/30$, $d_1 = 1, \dots, d_6 = 1$, $m_1 = 10.5, \dots, m_6 = 15.5$, $h_1^* = 3, \dots, h_6^* = 1$.

6. (a) -

(b) -

(c) „Wenigstens zwei Flaschen,“ wird an 76% der Verkaufstage erreicht